FACULTAD DE CIENCIAS

DEPARTAMENTO DE BIOLOGÍA

Gene editing in Hematopoietic Stem Cells: a potential therapeutic approach for Fanconi anemia

Memoria presentada por BEGOÑA DÍEZ CABEZAS licenciada en Biología, para optar al grado de doctor por la Universidad Autónoma de Madrid, con mención europea.

Directores de Tesis

Juan A Bueren Roncero

Paula Río Galdo

Madrid, 2015

CENTRO DE INVESTIGACIONES ENERGÉTICAS, MEDIO AMBIENTALES Y TECNOLÓGICAS (CIEMAT) – CENTRO DE INVESTIGACIONES BIOMÉDICAS EN RED DE ENFERMEDADES RARAS (CIBERER) – INSTITUTO DE INVESTIGACIÓN SANITARIA DE LA FUNDACIÓN JIMÉNEZ DÍAZ (IIS – FJD)

Paula Río Galdo, investigadora post doctoral de la División de Terapias Innovadoras del Sistema Hematopoyético del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, y **Juan Antonio Bueren Roncero**, jefe de la misma unidad, certifican que la memoria adjunta titulada *Gene editing in Hematopoietic Stem Cells: a potential therapeutic approach for Fanconi anemia* ha sido realizada por Begoña Díez Cabezas bajo la dirección conjunta de los que suscriben, y cumple las condiciones exigidas para optar al título de Doctor con mención europea por la Universidad Autónoma de Madrid.

Juan Antonio Bueren Roncero

Paula Río Galdo

El trabajo de investigación descrito en esta memoria ha sido realizado en la División de Terapias Innovadoras en el Sistema Hematopoyético del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) y, el Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER) y la Unidad Mixta de Terapias Avanzadas CIEMAT/Instituto de Investigación Sanitaria - Fundación Jiménez Díaz (IIS-FJD).

Para su ejecución, el trabajo de investigación realizado ha contado con la colaboración de los siguientes Programas de Investigación:

- Séptimo Programa Marco de la Comisión Europea (Proyecto PERSIST; Ref 222878).
- Ministerio de Economía y Competitividad (Proyectos SAF 2012-39834).
- Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III (RETICS-RD06/0010/0015; RD12/0019/0023).
- Dirección General de Investigación de la Comunidad de Madrid (Proyecto CellCAM; Ref S2012/BMD-2420).
- Programa de transferencia de tecnología en el campo de la terapia génica de la Fundación Botín.

Begoña Díez Cabezas ha disfrutado de una beca de Formación del Profesorado Universitario (FPU) del Ministerio de Educación, Cultura y Deporte (Ref AP2009-4413) y de un contrato del IIS-FJD.

A mis padres

Con estas líneas sólo quiero dejar constancia de que, pese a que en la portada sólo figura mi nombre, el trabajo que hay aquí es el resultado del esfuerzo, la paciencia, las ganas y la confianza de un montón de gente que ha estado a mi lado andando este camino. Podría llenar mil páginas con una sola palabra, GRACIAS, y aun así no reflejarían todo el agradecimiento que siento y os merecéis todos. ¡¡¡¡MILLONES DE GRACIAS A TODOS!!!!

Lo primero es agradecer por supuesto a mis directores de tesis, Juan y Paula. A Juan por haber confiado en mí sin conocerme y darme la oportunidad de formar parte de esta gran familia, pero sobre todo por haber aguantado tan estoicamente mis agobios en tema papeles, sobre todo durante el periodo de escritura y depósito. Paula, a ti te lo debo todo, eres la persona que más te has preocupado porque este trabajo saliera adelante, siempre motivándome aun cuando yo no era capaz de ver nada positivo (que sabes que es bastante a menudo). Tuviste confianza en mí desde el primer día, algo que aunque a veces se ha hecho duro, creo que ha sido uno de los pilares que ha sustentado este trabajo y por el cual hoy estoy orgullosísima. Gracias por tu guía, por tu ayuda, por tu comprensión y no sólo con las cosas del labo, gracias por hacer que esto hoy sea una realidad. ¡Es un placer trabajar contigo! Y además quiero que sepas que la mayor parte de mi crecimiento profesional te lo debo a ti, y por ello mil y una veces te daré siempre las gracias.

Los siguientes a los que quiero agradecerlo todo son los miembros de nuestro pequeño grupo dentro de la familia hematopoyética. La primera a Roci, por dejar que estropeara el nombre más molón del labo según Diego (Parola) dejándome poner la be al final (Parolabe), por haber tenido la paciencia de aguantarme y enseñarme a hacerlo todo pero sobre todo, tanto a ti como a Lara, por acogerme como lo hicisteis cuando llegué como una extraña. A Lara, que por supuesto si alguien se ha currado este trabajo tanto, o incluso diría que más que yo, has sido tú. Porque trabajar contigo siempre ha sido muy fácil y porque siempre has estado dispuesta a currar todo lo que hiciera falta. También quiero agradecerte no sólo la cantidad, sino la calidad del trabajo, ¡los resultados más bonitos son los tuyos! Gracias siempre por escucharme, por aguantarme, por intentar tranquilizarme y por consolarme en el día a día y también por dejarme compartir parte de tu vida. A Isa, que aunque estuviste poquito por aquí dejaste una huella muy grande, contagiándonos tu alegría y tu ilusión por este trabajo, que por otro lado es tantas veces tan ingrato, ¡se te echa de menos! Y por último a Fran, GRACIAS también con mayúsculas, no eres sólo un gran compañero e investigador y una de las personas más inteligentes y buenas que conozco, de hecho diría que ¡eres la persona más buena que he conocido en mi vida! Pero sobre todo por ser un gran amigo, preocuparte constantemente por mí y ayudarme en todo. Gracias de verdad, de corazón, a este pequeño grupo dentro de la gran familia de hemato, porque sin vosotros esto jamás habría salido adelante ni hubiéramos conseguido tanto.

Por supuesto a todo el labo, la división llamada hematopoyesis en la que entré, y la de terapias innovadoras en la que estoy ahora. ¡Venir a trabajar todos los días es un placer con compañeros como vosotros! Y además nunca podría imaginar otro lugar mejor en el que hacer la tesis, porque todo lo habéis hecho fácil y divertido. Para mi os habéis convertido todos y cada uno de vosotros en una parte importantísima de mi vida y os considero parte de mi familia. Gracias por aguantarme y soportar mi música y mis habilidades (o no) de canto, y por aguantar mis locuras todos los días, que no han sido pocas. Gracias por las comidas en el chino,

los concursos de tortillas, los desayunos para celebrar cualquier cosa, cenas de navidad y fiestas sorpresa de tesis.

A mis niñas, mis codoctorandas. A María, porque a pesar de no habernos conocido durante la carrera que estudiamos a la vez ni en el viaje de ecuador que hicimos juntas, eres no sólo una amiga y compañera que me ha entendido siempre y apoyado en casi todo (en la cantidad de asteriscos de significación que se deben poner no, pero no podíamos coincidir en todo ¿no?); eres un ejemplo de persona trabajadora y eficiente, sé que te lo han dicho muchas veces, pero ¡eres la organización en persona! Espero que durante estos años pasados juntas se me haya pegado algo de ti. Y además eres un ejemplo de superación y de saber enfrentarte a todo lo que la vida te depara; sobre todo te admiro por tu capacidad de enfrentarte y superar, o al menos saber vivir con todo, lo bueno y lo malo. A Fati, por haber compartido no sólo protocolos para transfectar o matar $CD34^{+}$ según se mire, sino opiniones sobre la vida, confidencias, cosas que jamás habría contado a nadie antes de conocerte. Para mí, reflejas perfectamente la parte pasional de vida, la que te hace disfrutar, y luchar por lo que realmente quieres. Te admiro también por saber decir las cosas, algo que ya sabes que a mí a veces me cuesta, y por ser defensora a ultranza de lo que quieres y valoras. A vosotras dos especialmente os doy las gracias por haberme aguantado más directamente que nadie, sobre todo durante este último periodo de escritura, en el que gracias a vosotras no me he vuelto loca de remate; pero también gracias por haberme aguantado y aconsejado siempre, porque todos los consejos que me habéis dado han sido siempre buenos (o por lo menos yo sólo recuerdo esos). A MJ, a pesar de que seas como un terremoto en el labo, eres una persona divertida y sin maldad con la que, al igual que con Fati, no sólo he compartido las mil y una maneras de matar a las células con las nucleasas, hemos compartido penas, pero sobre todo muchas alegrías y momentos inolvidables. Gracias por tu alegría contagiosa que me ha hecho sonreír tanto estos años. A Vic, gracias por el maravilloso diseño de colores de esta tesis, nadie habría podido hacerlo mejor que tú, consiguiendo que reflejara perfectamente mi personalidad. Gracias por todas las conversaciones alejadas de la ciencia, por darnos otra visión diferente y a veces diría más realista de las cosas que hacemos día a día. Y aunque codoctoranda no y a pesar de que siga desarrollando su carrera de niño repu en London, por supuesto a Diego. Gracias por preocuparte siempre por mí, y además por seguir haciéndolo desde la distancia. Por ser compañeros no sólo de beca, sino de un montón de cosas más: aficiones, maneras de ver y disfrutar la vida, por los momentos de locura y de organización de performances varias, pero muy especialmente por tener siempre voluntad para integrar en vez de separar algo que de verdad admiro profundamente de ti. A todos vosotros gracias por vuestra ayuda en el día a día en todo, en el labo con experimentos, pero también gracias por instaurar tradiciones como el Adeste Fideles el día antes de irnos de vacaciones de Navidad, los momentos del google translator, las bromas, porque habéis hecho que todos los días pensase lo afortunada que soy de haber compartido mi día a día con vosotros.

También agradecer a todos los ya doctores que he tenido la fortuna de conocer y poder aprender con y de ellos: Sandra, Elenita, Zita, María peque, Álvaro, y especialmente a Javi, por siempre darme buenos consejos, pero sobre todo por tus abrazos que siempre me han venido tan bien y que desde que te fuiste a Boston tanto echo de menos.

A los pequeños, Sergio, Cristina, Victoria epiteliana, Sara y Carlos. Gracias por todos los buenos momentos compartidos, porque llegasteis en un momento en el que me ayudasteis mucho, por haberme dejado enseñaros aunque sea un poquito y por a pesar de la diferencia de edad que nos separa haber querido salir, disfrutar y compartir tantas cosas dentro y fuera del labo.

A citometría team, Rebe y Oma. Vosotras dos habéis soportado mi impaciencia pasando miles de tubos, ¡cuántas veces os habré dicho que por qué en el citómetro no ponen un botón de super high!; mis charlas de todo y nada durante esos experimentos de ratones interminables en los que ¡ni siquiera podía centrifugar todos los tubos de una vez porque no cabían en la centri! Porque las dos sois unas currantas como he visto pocas en mi vida y también por convertiros en amigas y a ti Rebe por dejarme formar parte de momentos tan importantes de tu vida.

A todos los técnicos, Raquel, Mari Luz, Isra, Laura, Montse y Miriam. Por haberme enseñado y ayudado cuando ha hecho falta, y además hacerlo siempre con una sonrisa y con buen humor. Y también a todos los técnicos que han pasado y aprendido o están aprendiendo con nosotros estos años, especialmente a Javi; de todos siempre he podido aprender algo y espero haber contribuido a que aprendáis de alguna manera.

A la bati, Oscar, Mercedes, Susana, Elena, María grande y Rosa. Por vuestros buenos consejos a la hora de hacer experimentos, por contarme todos los pequeños trucos para que todo salga mejor y ayudarme con todos vuestros conocimientos científicos. Pero también por todas las horas de comida compartidas y las conversaciones absurdas que ha habido en ellas. Rosa, quiero además agradecerte el que hayas sido un poco mi madre estos años (y lo digo en el buen sentido), y te hayas preocupado tanto por mí y me hayas escuchado tantas veces.

A los compis de despacho de Paula, Maru, Jaco y Marina; por haber soportado que os quitáramos las sillas todos los lunes para nuestras reuniones, pero sobre todo por los consejos que me habéis dado durante toda la tesis, aunque no los pidiese ¿verdad?. Y también gracias por todas las horas de comida compartidas en las que tan buenos ratos hemos pasado.

A los co-jefes: José Carlos y Guille. Gracias por toda vuestra ayuda y organización. Guille decirte que sin ti este laboratorio no sería el mismo.

A toda la gente que ha estado conmigo aunque haya sido brevemente, José, Clara, Raquel (alias Famous), Miguel y Estrella. Gracias por vuestra alegría y ganas, por haber compartido también tantos buenos momentos y tantas confidencias. Pero especialmente a Clara, porque te has convertido en amiga a pesar de tus bombas de humo.

Thanks also to all Naldini's laboratory, for helping me during my stay there, give me the opportunity to learn everything I learnt and to work in one of the most important laboratories in gene therapy all over the world; but especially to PIM and Giulia. Both of you teach me everything that I know about how not to kill $CD34^+$ cells, help me every day that I was there, and if today I have positive results in my gene targeting experiments is because of you. I will be extremely grateful to you.

A pedidos: Mamen, Sergio, Aurora y Sole. Chicos gracias pese a que os haya vuelto locos de vez en cuando con los pedidos, con lo de que se me ha acabado no sé enzima o kit y lo necesito

para ya, siempre me habéis salvado la vida. Sin vuestra organización, nuestro trabajo sería imposible.

A Miguel por toda la cantidad de papeleo resuelto, ¡solo recuerdo el lio de los tickets presentados durante mis tres meses de estancia!, por compartir o por lo menos estar más cerca que nadie en gustos musicales, pero sobre todo por resolver siempre tantos marrones.

A María José, Freddy, Rocío y Rosita porque vosotros también sois parte del laboratorio y sin vosotros sí que todo sería un caos inimaginable.

A todos los miembros del animalario, pero especialmente a Miguel Ángel, por cuidar a nuestros ratones como si fueran suyos y avisarnos siempre de cualquier cosita extraña que notase.

A Norman, por su inestimable ayuda con el inglés sobre todo durante la escritura de este libro.

A epitelios, tanto grandes como pequeños, porque siempre me habéis ayudado cuando os he necesitado.

Y por supuesto a las familias y pacientes con anemia de Fanconi, ellos son el porqué de este trabajo, algo que considero absolutamente fundamental aunque en el día a día a veces se nos olvide.

¡Y esto era sólo del CIEMAT!

También me quiero acordar de toda la gente que aunque no tan directamente, han hecho posible que al final llegase a formar parte de esta familia. Lo primero a mis antiguos jefes: Jeremy Rhodes, Jaime Gosalvez y María Montoya, todos me disteis la oportunidad de empezar y continuar en ciencia y guiasteis mis pasos hasta aquí. A mis primeros compis de laboratorio, en genética de la UAM, pero sobre todo a mis titis, Clara y Laura, ahora tan lejos de aquí pero tan felices las dos, y no sabéis lo que me alegro por ello. También por supuesto a mis compis del CNIC: Raquel y Mariano por enseñarme todo lo que sé de citometría pero sobre todo ¡por enseñarme a entenderla! A Iñaki y Montse por ser un apoyo y ayuda constante en el laboratorio, por todos los buenos ratos vividos en cultivos con nuestros compis de vectores Raúl y Aída. A Elvira por enseñarme tanto y tan bien de lo que se puede hacer en microscopía. A Inma por enseñarme a clonar. Pero sobre todo a Ligos porque además de enseñarme un montón de cosas en el labo, fuiste tú el que me aconsejaste venirme aquí, y por ello te estaré agradecida toda la vida.

Y ya por último y entrando en el terreno personal, a todos los que aun no entendiendo nada de mi trabajo, me habéis apoyado toda o casi toda mi vida, y que de manera todavía más indirecta, también habéis contribuido a esto.

A mis amigos, los que lleváis conmigo toda la vida Aída y Pedro, por seguir ahí pese a todo y pese a todos compartiéndolo todo. Especialmente a Aída porque a pesar de que a veces no lo ponga nada fácil sigues a mi lado. A los que compartisteis conmigo la carrera Víctor, Gema, Carmen y David, por hacer que mi paso por la UAM fuera tan genial e inolvidable y haberos convertido en personas tan importantes. Pero especialmente a Carmen por no sólo haber

compartido carrera sino también labo en el CNIC y penas durante nuestro paso por él; y a David, por ayudarme tanto no sólo personal sino también profesionalmente (has hecho la portada de tesis más bonita del mundo ¡y eso es sólo un ejemplo!), y por ser una persona a la que admiro y tengo como referencia para muchísimas cosas. A los que fuisteis añadidos por diversas circunstancias y casualidades pero que sois también miembros de mi familia: Bego, Dani, Toño, Lolo, Javi, José, Paco, Álvaro y Marisa; por darme todo vuestro cariño siempre, por entenderme, por todos los buenos momentos y vuestro apoyo. Pero sobre todo quiero agradecerte a ti Bego todo lo que has hecho por mí y todo el apoyo que siempre me has dado, en especial estos últimos meses de estrés absoluto. A Elena, Marta e Irene, porque aunque os conocí por ser la novia de alguno de mis amigos hoy tengo el privilegio también de contaros ente mis amigas; pero especialmente a Marta por todo el apoyo que me has dado durante esta última etapa de la tesis desde tu experiencia personal. Gracias de verdad a todos porque aunque haya sido de manera más indirecta que nadie también habéis hecho este día posible.

También quiero agradecer a Jorge, Diana y Tristán, que a pesar de no tener porqué también habéis permanecido a mi lado.

A la familia Cid Velasco, por haber sido mi familia durante muchos años, por el cariño y el apoyo que siempre me habéis brindado.

Y por supuesto y por último aunque no menos importante, a toda mi familia, que aunque no seamos la típica familia que está junta todo el día, siempre estamos en los momentos más importantes, tanto buenos, en los que siempre es más fácil estar, pero sobre todo en los malos, que es cuando realmente más se necesita.

Pero sobre todo agradecer a mis padres y a mi hermano. Charly, aunque no nos llamemos tanto como papá querría, siempre me has apoyado en todo (o casi) y has estado conmigo cuando más te he necesitado consolándome. Y por supuesto gracias por todas las risas, tonterías y el cariño que siempre me has dado, aunque la mayor parte de las veces haya sido a base de estrujarme. Y por último a mis padres, por habérmelo dado todo y por seguir dándomelo, por entenderme o intentarlo, por apoyarme, por aconsejarme, porque si hoy soy quien soy, pero sobre todo, si tengo las cosas buenas que tengo, os las debo a vosotros. Por aguantarme aunque a veces lo ponga tan difícil, y porque a pesar de haberos decepcionado alguna que otra vez me seguís queriendo tan incondicionalmente. Gracias a los tres por haber estado y siempre estar.

Lo que consigues con el logro de tus metas no es tan importante como en lo que te conviertes al conseguirlas

Henry David Thoreau

Si caminas solo irás más rápido pero si caminas acompañado llegarás más lejos

Proverbio chino

RESUMEN

SUMMARY

La terapia génica constituye hoy en día una alternativa segura y eficaz para el tratamiento de determinadas enfermedades monogénicas que afectan al sistema hematopoyético. Sin embargo, puesto que el riesgo de mutagénesis insercional debido al uso de vectores integrativos no se puede descartar completamente, la terapia génica de edición se ha propuesto como una alternativa más segura, ya que la inserción del gen terapéutico está dirigida a un *locus* específico del genoma. Las aproximaciones de edición génica se basan en el uso de nucleasas específicas que generan roturas de doble cadena en un sitio específico del genoma, mejorando así la eficacia de la recombinación homóloga con construcciones donadoras que albergan el gen de interés flanqueado por los brazos de homología correspondientes. En este estudio se han optimizado las condiciones para editar líneas celulares linfoblastoides (LCLs) y células madre hematopoyéticas (CMHs), inicialmente de donantes sanos, con el objetivo final de corregir mediante terapia génica de edición las progenitores hematopoyéticos de pacientes con anemia de Fanconi del subtipo A (AF-A).

En particular, hemos establecido un método eficaz para editar LCLs y CMHs en un sitio seguro en el genoma, el locus *AAVS1*. Nuestro protocolo se basa en la transducción de estas células con vectores lentivirales defectivos en integración que portan un donador con el gen de interés, seguido de la nucleofección de estas células con nucleasas de dedos de zinc (ZFNs) utilizadas como ARNm. Utilizando un vector donador control que lleva el gen marcador *GFP* hemos logrado una eficiencia media de corrección génica dirigida en células $CD34^+$ de donantes sanos del 9,43%. Por otra parte, hemos confirmado que la edición génica también es eficaz en CMHs multipotentes con capacidad de repoblación a largo plazo en ratones inmunodeficientes. Para mejorar la eficacia de la edición génica, hemos investigado la posibilidad de utilizar nanopartículas de oro, que incrementaron la eficiencia de transducción de vectores lentivirales integrativos y no integrativos en las células $CD34^+$. Sin embargo, este incremento no dio lugar a una mejora en la edición génica, debido a la toxicidad asociada a la nucleofección de las células tratadas con las nanopartículas. A continuación, mediante el uso de un vector donador terapéutico, portador del gen *FANCA*, hemos demostrado la corrección del fenotipo de LCLs de pacientes con AF-A mediante edición génica, que confirmamos por la formación de focos de reparación de FANCD2 y la reversión de la sensibilidad de estas células frente a un agente entrecruzante del ADN, tal como mitomicina C (MMC). Para mejorar la eficacia de edición génica en células hematopoyéticas de pacientes AF-A, también investigamos el efecto mediado por la inhibición transitoria de la anti-recombinasa PARI. Aunque la inhibición de PARI aumentó los focos de reparación de RAD51, no observamos aumento significativo en la eficiencia de la edición génica. Por último, hemos demostrado que nuestro protocolo de edición génica también permite, aunque con menor eficacia, la inserción del gen terapéutico en progenitores hematopoyéticos de pacientes con AF-A, permitiendo así la reversión parcial de su sensibilidad a MMC.

Nuestro estudio demuestra por primera vez que la terapia génica de edición en el *locus* seguro *AAVS1* es factible en las células hematopoyéticas, incluyendo progenitores hematopoyéticos primarios, de pacientes con anemia de Fanconi del subtipo A, abriendo nuevas perspectivas para la terapia génica dirigida de esta enfermedad, así como de otras enfermedades monogénicas del sistema hematopoyético.

SUMMARY

Gene therapy nowadays constitutes a safe and efficient treatment for a number of monogenic diseases affecting the hematopoietic system. Risks of insertional mutagenesis derived from the use of integrative vectors cannot, however, be completely excluded. Therefore, gene targeting has been proposed as a safer alternative, since the insertion of the therapeutic gene is driven to a specific locus in the genome. Gene targeting approaches are based on the use of specific nucleases which generate double strand breaks (DSBs) in a specific site of the genome, markedly enhancing the efficacy of homologous recombination (HR) with donor constructs harboring the gene of interest flanked by the corresponding homology arms. In this study we have optimized the conditions to target human lymphoblastic cell lines (LCLs) and also hematopoietic stem cells (HSCs) from healthy donors, with the final aim of correcting by gene editing the hematopoietic progenitor cells from Fanconi anemia subtype A (FA-A) patients.

In particular, we have established a robust method to target both LCLs and HSCs in a safe harbor site in the genome, the *AAVS1* locus. Our approach is based on the transduction of these cells with integrase-defective lentiviral vectors carrying a donor with the gene of interest, followed by the nucleofection of these cells with zinc finger nucleases used as mRNA. Using a control donor vector carrying the *GFP* reporter gene we have obtained, on average, 9.43% gene targeting efficiency in cord blood $CD34^+$ cells from healthy donors. Moreover, we confirmed that gene targeting was also efficient in HSCs with long term and multipotent repopulation capacity, as demonstrated by transplants into immunodeficient mice. To improve the gene targeting efficiency, we investigated the feasibility of using gold nanoparticles, which were shown to improve the transduction efficiency of integrase-defective and competent lentiviral vectors in HSCs. This increment, however, did not lead to a higher gene targeting efficiency, due to the toxicity associated with the nucleofection of cells treated with these nanoparticles. In our next step, we moved from healthy donor HSCs to FA hematopoietic cells. Using a therapeutic donor vector carrying the *FANCA* gene, we demonstrated that gene targeting can correct the phenotype in a FA-A LCL. This was deduced from the restoration of FANCD2 foci formation and the reversion of the sensitivity of FA-A cells to interstrand cross linkers, such as mitomycin C (MMC). To improve the gene targeting efficiency in FA-A hematopoietic cells, we also investigated the effects mediated by the transient inhibition of anti-recombinase PARI. Although the inhibition of PARI increased RAD51 foci, no significant increase of homology directed repair efficiency was observed. In a final set of experiments we demonstrated that our gene targeting approach has also taken place in hematopoietic progenitor cells from FA-A patients, leading to a partial reversion in their hyper-sensitivity to MMC.

Our study demonstrates for the first time that gene targeting in the *AAVS1* safe harbor locus is feasible in hematopoietic cells from Fanconi anemia-A patients, opening up new perspectives for the future gene therapy of this and other monogenic diseases of the hematopoietic system.

INDEX

INDEX

ABBREVIATIONS

53BP1	p53 binding protein 1
aa	Amino acid
AAV	Adeno-associated vector
AAVS1	Adeno-associated virus integration site 1
ADA-SCID	Adenosine deaminase deficiency-severe combined immunodeficiency
AdV	Adenoviral vector
AF	Anemia de Fanconi
AF-A	Anemia de Fanconi del subtipo A
ALD	Adrenoleukodystrophy
AML	Acute myeloid leukemia
APC	Allophycocyanin
APLF	Aprataxin and PNKP like factor
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and RAD3-related
AuNP	Gold nanoparticle
BFU-Es	Burst forming unit-erythroid
BLM	Bloom syndrome, RecQ helicase-like
BM	Bone marrow
BMF	Bone marrow failure
B-NHEJ	Alternative/backup non-homologous end joining
BRCA1	Breast cancer susceptibility 1
BRCA2	Breast cancer susceptibility 2
BTR	BLM-Topoisomerase IIIα-RMI1-RMI2 complex
BV	Baculoviral vector
Cas	CRISPR associated
CCR5	Chemokine (C-C motif) receptor 5
CFC	Colony forming cell
CFUs-GMs	Colony forming units-granulocyte/macrophage
CGD	Chronic granulomatous disease
CHEK1	Checkpoint kinase 1
CMH	Célula madre hematopoyética
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSA	Cockayne syndrome type A
CSB	Cockayne syndrome type B
CtIP	C-terminal binding protein interacting protein
CXCR4	Chemokine receptor 4
DDT	DNA damage tolerance
DEB	Diepoxybutane
DNA2	DNA replication helicase/nuclease 2
D-NHEJ	DNA-PKcs dependent or canonical non-homologous end joining
dNTPs	Deoxynucleotides
DSB	Double strand break
DSBR	Double strand break repair
dtp	Days post-transplant
EPO	Erythropoietin
Exo1	Exonuclease 1
FA	Fanconi anemia
FA-A	Fanconi anemia complementation group A
FAAP	Fanconi associated-protein
FAN1	Fanconi associated-protein 1
FANCA	Fanconi anemia subtype A
Flt3	FMS-like tyrosine kinase 3 ligand
Fw	Forward
G-CSF	Granulocyte-colony stimulating factor

gDNA	genomic DNA
GG-NER	Global genome nucleotide excision repair
gRNA	Guide RNA
GT	Gene therapy
GVHD	Graft versus host disease
HA L	Left homology arm
HA R	Right homology arm
HA	Homology arm
hAlb	Human albumin
HD	Healthy donor
HDR	Homology-directed repair
HPV	Human papillomavirus
HR	Homologous recombination
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
HSPC	Hematopoietic stem and progenitor cells
ICL	Interstrand-crosslinks
ICLV	Integrase-competent lentiviral vector
ID2	FANCD2/FANCI complex
IDLV	Integrase-defective lentiviral vector
IMDM	Iscove's Modified Dulbecco's Medium
INDELs	Insertions and deletions
iPSC	Induced pluripotent stem cell
iRNA	Interference RNA
kb	Kilobase
LCL	Lymphoblastoid cell line
LSK	Lin^-, $sca1^+$, $ckit^+$ phenotype
LV	Lentiviral vector
MDS	Myelodysplastic syndrome
MGN	Meganuclease
MHF1	Histone fold protein 1
MHF2	Histone fold protein 2
MLD	Metachromatic leukodystrophy
MMC	Mitomycin C
MOI	Multiplicity of infection
MOPs	3-(*N*-morpholino)propanesulfonic acid
mPB	Mobilized peripheral blood
MRN	Mre11-RAD50-Nbs1 complex
NBS	Nijmegen breakage syndrome
NBS1	Nijmegen breakage syndrome 1 protein
NEEA	Non-essential aminoacids
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NK	Natural killer
NLS	Nuclear localization signal
nm	nanometer
NP	Nanoparticle
NSG	NOD.Cg-*Prkdcscid IL2rgtm1Wjl*/SzJ
NSG-SGM3	NOD.Cg-*Prkdcscid*Tg (hSCF/hGM-CSF/hIL3)
P/S	P**e**nicillin/Streptomycin
PAM	Protospacer-adjacent motif
PARI	PCNA-associated recombination inhibitor
PARP-1	Poly (ADP-ribose) polymerase 1
PB	Peripheral blood
PBS site	Primer binding site
PCNA	proliferating cell nuclear antigen
PE	Phycoerythrin

PGK	Phosphoglycerate kinase
PNKP	Polynucleotide kinase 3'-phosphatase
PPP1R12C	Phosphatase I regulatory inhibitor subunit 12C
PTD	Protein transduction domain
Puro	Puromycin
Q-PCR	Quantitative PCR
RH	Relative humidity
ROS	Reactive oxygen species
RPA	Replication protein A
RPMI	Roswell Park Memorial Institute
RQ	Relative quantity
Rv	Reverse
RV	Retroviral vector
RVD	Repeat-variable di-residue
SA	Splicing acceptor site
SCC	Squamous cell carcinomas
SCD	Sickle-cell disease
SCF	Stem cell factor
SD	Standard deviation
SDSA	Synthesis-dependent strand annealing
SF	Steel factor
SIN γ-RV	Self-inactivating gamma retroviral vector
SIN LV	Self-inactivating lentiviral vector
ssDNA	Single- stranded deoxyribonucleic acid
ssRNA	Single-stranded RNA
TALEN	Transcription activator-like effector nucleases
TAR	Thrombocytopenia absent radius
TBS-T	TBS Tween buffer
TC-NER	Transcription-coupled nucleotide excision repair
TE	Tris-EDTA
TEV	Tobacco etch virus
TLS	Translesion-synthesis
TNF-α	Tumor necrosis factor α
TPO	Thrombopoietin
tracrRNA	trans-activating crRNA
UBE2T	Ubiquitin-conjugating enzyme E2T
UCB	Umbilical cord blood
UTR	Untranslated region
VCN	Vector copy number
VLDI	Vector lentiviral defectivo en integración
vp	viral particles
WB	Western blot
WRN	Werner syndrome, RecQ helicase-like
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4
X-SCID	X-linked chronic granulomatous disease
ZF	Zinc-finger
ZFN	Zinc finger nuclease
γ-RV	Retroviral vector

INTRODUCTION

1. Conventional gene therapy

Gene therapy (GT) consists in the delivery of genetic material into patient's cells to treat a disease. It can be achieved either by the introduction, the substitution or the alteration of genes. This approach is especially relevant for the treatment of monogenic diseases in which the therapeutic gene can correct the genetic defect[1-3]. Therapeutic genes can be introduced in the cells using different mechanisms such as retroviral vectors (RVs) and lentiviral vectors (LVs) that will integrate into the genome and will allow the long term expression of the transgene. Classically, viral vectors are widely used for this purpose because of their higher efficiency; however, there are also non-viral methods to introduce the genetic material inside the cells, which include transposons, liposomes, inorganic nanoparticles and others[4]. The outcome of gene therapy can be improved if the correction of mutated cells confers them a proliferative or survival advantage over non-modified ones. This is the case of Fanconi anemia (FA) patients with somatic mosaicism, in which the reversion of the pathogenic mutation mediates a proliferation advantage of corrected cells, representing a model of natural gene therapy (see section 2.5. of the introduction)[5].

According to the modality of treatment, GT can be classified as *in vivo* or *ex vivo* GT. The *in vivo* modality consists in the introduction of a therapeutic gene directly into the patient by the inoculation of the vector that carries the transgene. On the other hand, the *ex vivo* GT is based on the *in vitro* modification of cells previously collected from the patient, which are then reinfused back (Figure 1). Moreover, and according to the integration nature of the transgene in the cell genome, we can distinguish two types of GT: Conventional GT, in which the integration of the new genetic material is not directed to specific loci; and the Targeted GT in which the integration is site-specific. Both of them will be discussed in the next sections.

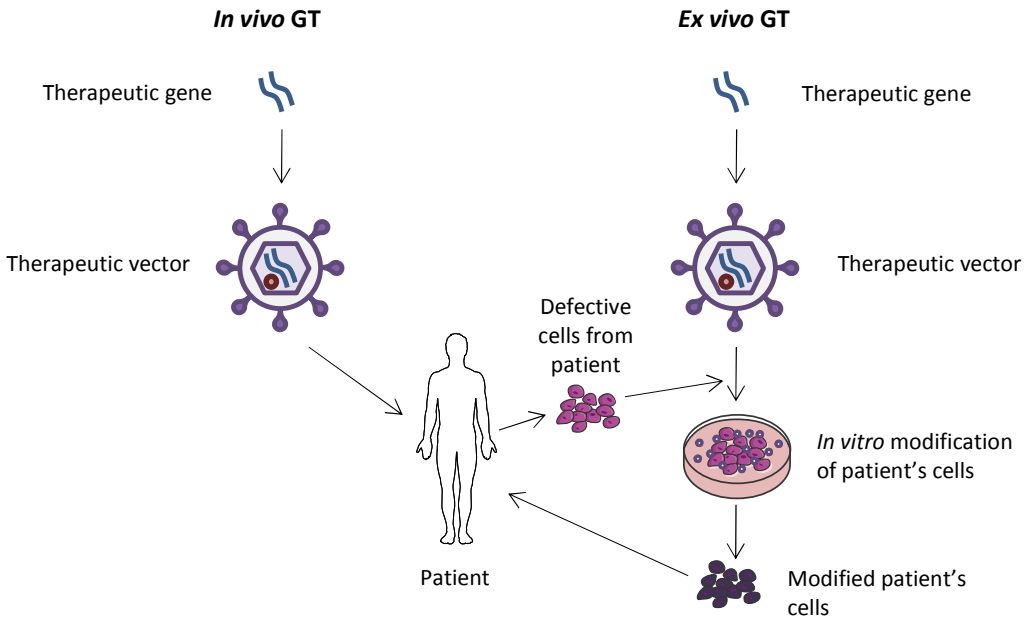


Figure 1: Schematic representation of in vivo versus ex vivo gene therapy*. In vivo gene therapy consists in the direct inoculation of the vector carrying the therapeutic gene in the patient. Ex vivo gene therapy*

consists in the isolation of target cells from the patient, followed by the in vitro treatment with the therapeutic vector and finally the reinfusion of these modified cells in the patient. GT: Gene Therapy.

1.1. Conventional *ex vivo* gene therapy

As previously mentioned, conventional GT is characterized by the integration of a therapeutic gene in a non-targeted way. Viral vectors are widely used to conduct this kind of GT. Conventional *ex vivo* GT is the strategy most widely used in hematopoietic stem cell (HSC) gene therapy, as it reduces the immunological response[3]. Moreover, HSCs are the principal target cells for this approach due to the feasibility to isolate them *ex vivo*, genetically modify and transplant them again by intravenous infusion. Once genetically modified HSCs engraft into the bone marrow (BM) of the patient; these cells can produce new and healthy blood cells of all lineages for the whole life of the patient. Autologous HSC GT has been successfully used in different clinical trials during the last years (Table 1). Monogenic hematological disorders are currently the main indication, and these include primary immunodeficiencies: such as adenosine deaminase deficiency-severe combined immunodeficiency (ADA-SCID), X-linked SCID (X-SCID), chronic granulomatous disease (CGD) and Wiskott-Aldrich syndrome (WAS); erythroid diseases like β-thalassemia and sickle cell disease (SCD) and bone marrow failure (BMF) syndromes like FA. Storage disorders like metachromatic leukodystrophy (MLD) and adrenoleukodystrophy (ALD) have been also treated by autologous HSC GT due to the ability of HSC to deliver the therapeutic enzyme to the nervous system.

Disease	Target cells	Transduced cells	Vector	Treated patients	Efficacy	Adverse effects
ADA-SCID[6-8]	Lymphoid cells and natural killer cells (NK)	HSC	γ-RV	42	Long-term cure comparable to HSCT	No
X-SCID[9-11]	Lymphoid cells and NK	HSC	γ-RV	25	Long-term cure comparable to HSCT	Yes
X-SCID[12-14]	Lymphoid cells and NK	HSC	SIN γ-RV	9 (on going)	Recovery of PB T cells that were functional and eliminated the infections	No
CGD[15]	Neutrophils	HSC	γ-RV	23	Transient or no clinical benefit. Long-term efficacy in the three cases of clonal expansion followed by inactivation of the transgene	Yes
WAS[16,17]	Lymphoid cells, NK and platelets	HSC	γ-RV	10	Clear benefit at 2 year follow-up	Yes

Disease	Target cells	Transduced cells	Vector	Treated patients	Efficacy	Adverse effects
WAS[18-20]	Lymphoid cells, NK and platelets	HSC	SIN LV	4 (on going)	Improvement in platelet counts, immune functions and clinical scores	No
β-thalassemia[21]	Erythroid cells	HSC	SIN LV	4	Therapeutic benefit. Transfusion independence	No
ALD[22]	Macrophages and microglia	HSC	SIN LV	4	Clear therapeutic benefit comparable to HSCT	No
MLD[23]	Macrophages and microglia	HSC	SIN LV	9 (on going)	Stopped the manifestation or progression of the disease beyond the predicted age of symptom onset	No

Table 1: Ex vivo HSC GT clinical trials. *NK: natural killer. γ-RV: gamma-retroviral vectors. HSCT: hematopoietic stem cell transplantation. SIN LV: self-inactivating lentiviral vectors. SIN γ-RV: self-inactivating gamma retroviral vectors. PB: peripheral blood. Modified from Naldini*[2] *and Gosh et al*[1].

The viral vectors most widely used for hematopoietic GT are γ-retroviral (γ-RV) and lentiviral vectors (LVs). Both belong to the same family of virus (*Retroviridae*) that possesses RNA based genome and present the same basic structure, 5' and 3' ends which flank the three viral genes (Figure 2):

- *Gag* gene encodes for capsid, matrix and nucleocapside proteins.
- *Pol* gene encodes the reverse transcriptase, the integrase and the protease.
- *Env* gene encodes the proteins for the viral envelope.

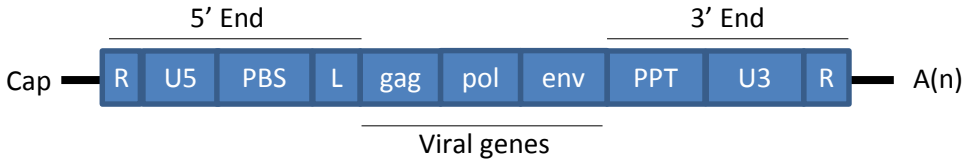


Figure 2: Retroviral RNA genome. *Retroviral genome consists in two single stranded RNA molecules with a Cap structure in the 5' end and a poly-adenylated tail in the 3' end. 5' and 3' ends contain some regulatory regions of the virus which flank the three viral genes: gag, pol and env.*

The viral cycle goes from single stranded RNA (ssRNA) through a dsDNA intermediate that is able to integrate into the host genome, allowing this proviral DNA to replicate together with the host DNA. The main differences between LVs and γ-RVs are the followings:

- The genome of the LVs is more complex than γ-RVs one, as it presents accessory proteins that are involved in the regulation of the viral cell cycle.

- LVs can transduce non-diving cells.
- LVs present a safer integration profile than γ-RVs, as they integrate preferentially inside transcription units of actively transcribed genes, in contrast to γ-RVs that presents a preferential integration in regions close to the transcription start sites of genes and CpG islands[24,25].

The major problem derived from the use of these vectors is the genotoxicity due to insertional mutagenesis, which has been clearly demonstrated with γ-RVs in different clinical trials, but not with LVs. Insertional mutagenesis is derived from the enhancer activation of the neighboring genes by viral LTRs, which are formed by reverse transcription of retroviral RNA at either end of the proviral DNA. LTRs are segmented into U3, R and U5 regions of the 5' and 3' ends of retroviral RNA genome (Figure 2) and participate in the integration of the provirus into the host genome. Once the provirus has been integrated, the LTR on the 5' end serves as the promoter of the entire retroviral genome and is the one responsible for the transactivation of neighboring genes. Several cases of hematological malignancies were described in the first γ-RV-based GT clinical trials: T-cell acute lymphoblastic leukemia cases in SCID-X1 and WAS patients have been associated with LTR-mediated up-regulation of a proto-oncogene (*LMO2*[26,27]), while clonal dominance and myelodysplastic syndromes (MDSs) described in γ-RV-treated CGD patients occurs by up-regulation of myeloproliferative genes like *MECOM* and *PRDM16*[15]. In order to limit the trans-activation potential of these vectors and to increase their safety some modifications have been performed. One of the most important ones consists in the inactivation of the U3 region (located in the 3' end of the viral genome) leading to the generation of self-inactivating (SIN) vectors and the use of more physiological promoters. Now is well proven that, by deleting the viral LTR, the risk of insertional mutagenesis through enhancer activation of the neighboring genes is significantly reduced[1]. Even more, the integration pattern of LVs is much safer and no adverse events have been detected in patients treated with these LVs with a follow up of 9 years. In spite of all these advantages, some studies have reported that LV integration can lead to the production of aberrantly spliced transcripts and deregulated gene expression in the integrated genes[28-30], although this phenomenon has not been observed in patients. Furthermore, it has been also described a partial clonal dominance in a LV-treated β-thalassemia patient as a consequence of increased expression of a truncated HMGA2 mRNA that escaped regulation by a miRNA directed to the 3' end of the full-length mRNA. However, this dominance was confined to erythroid cells, did not result in leukemic progression as in some γ-RV based cases and the clone stabilized 5 years post-gene therapy[21]. Nowadays, GT using LV vectors has become a reliable therapeutic option to treat some monogenic diseases with no adverse effects described up to now.

2. Directed/targeted gene therapy

Directed gene therapy or targeted gene therapy is characterized by the integration of the genetic material in a specific site of the genome. Using gene targeting approaches it is possible to interrupt, insert or replace a specific sequence in the genome in a determined locus and in a specific way. Directed GT increases the safety of GT, as the modification of the genome only

occurs in a specific region, avoiding thus the risk of insertional oncogenesis or deregulation of gene expression.

Nucleases are the tools selected to produce a double strand break (DSB) in a specific site in the genome and will be discussed in the section 2.3. of the introduction. These DSBs generated by nucleases are repaired by two endogenous DNA repair mechanisms: homologous recombination (HR) or non-homologous end joining (NHEJ); which will be explained in detail in the section 2.1. of the introduction. HR occurs at a very low frequency (in 1 out of 10^6-10^8 cells[31,32]) in mammalian cells and is only active in S-G2-phases of the cell cycle while NHEJ is the preferential pathway to repair this kind of lesions and is active during all the cell cycle. So, the cell cycle status is a crucial parameter to consider in gene targeting strategies[33]. The repair of these DSBs by NHEJ leads to stable gene disruption or to somatic knockout of a gene if both alleles are targeted. However, if the repair of DSBs generated by nucleases occurs through HR, it can cause insertions or substitutions in the endogenous cellular DNA through a process called homology-directed repair (HDR).

2.1. Endogenous mechanisms of double strand break DNA repair

Double strand breaks are, together with interstrand-crosslinks (ICLs), one of the most deleterious lesions that can occur in the DNA. DSBs can be produced by endogenous sources like reactive oxygen species, cellular metabolism, and replication associated errors; or by exogenous sources including ionizing radiation or chemotherapeutic agents. Unrepaired or misrepaired DSBs can result in senescence, induced apoptosis or chromosomal aberrations, including translocations and deletions that can finally produce cancer[34]. The two main pathways involved in DSB repair are NHEJ and HR[35].

2.1.1. Non-homologous end joining

Non-homologous end joining is an error prone mechanism that is preferentially used for the cell to repair DSBs (one thousand fold more frequent than HR[36]). This pathway is active during the different cell cycle phases but predominantly acts during G0 and G1 phases[37]. NHEJ implies the ligation of both DNA ends with minimal end processing[38] and usually generates small insertions or deletions called INDELs. NHEJ pathway can be subdivided in DNA-PKcs dependent/canonical NHEJ (D-NHEJ) and alternative/backup NHEJ (B-NHEJ) (Figure 3).

Canonical NHEJ (D-NHEJ) is produced when the heterodimer Ku70/Ku80 binds to DNA ends. This binding prevents nucleolitic processing of DNA ends which are required for initiation of other DSB repair pathways. The Ku heterodimer recruits DNA-PKc, which is the catalytic subunit of a DNA dependent kinase. The binding and dimerization of DNA-PKcs immobilizes the two DNA ends and thus facilitates the rejoining reaction[39]. Then, a limited end-processing by nucleolitic enzymes such as Artemis, PNKP, APLF, WRN, aprataxin or DNA polymerases μ and λ is required to generate ligatable ends[40]. Finally, the DNA ligase IV and XRCC4 ligate the two DNA ends to complete the repair of the damaged DNA[41].

Alternative NHEJ (B-NHEJ) is more prone to generate mutations than the canonical pathway, and thus frequently generates translocations and other genomic rearrangements. It is a backup pathway that is activated after failure of D-NHEJ[42,43] and is much slower than D-NHEJ.

This pathway benefits from microhomology at the break sites, which is more probably found if the DNA ends become resected, although microhomology is not strictly required. The major protein implicated in B-NHEJ is PARP-1, which may effectively compete with Ku heterodimer for DNA end-binding[44]. MRN complex (formed by MRE11- NBS1-RAD50) and CtIP are proteins involved in the DNA end-resection during HR and also facilitates B-NHEJ[45,46]. It is hypothesized that PARP-1 recognizes the DNA ends, the MRN complex and CtIP proteins resect them and then ligase III is recruited to the DSB to restore DNA integrity[47]. Some other proteins such as WRN has been reported to form a complex with ligase III, which activates the ligation process[48].

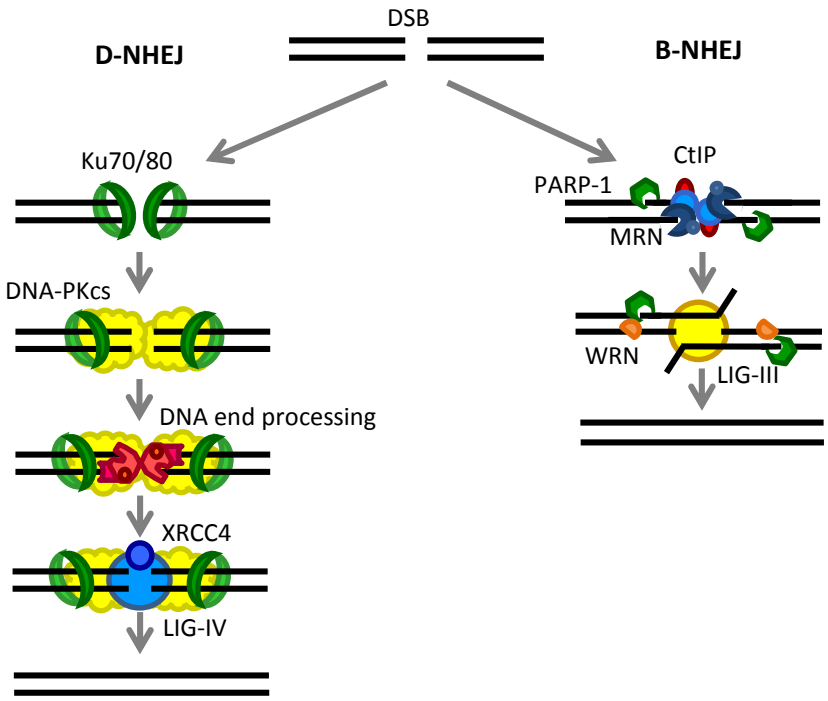


Figure 3: Non-homologous end joining (NHEJ) pathway *to repair double strand breaks (DSBs) through all cell cycle independently of the presence of homologous template. DSB could be repaired by D-NHEJ: DNA-PKcs dependent/canonical NHEJ in which the heterodimer Ku70/80 binds to DNA ends and recruits DNA-PKcs that immobilizes the two DNA ends. After a limited processing of these DNA ends, DNA ligase IV (LIG-IV) and XRCC4 ligate both DNA ends. DSB could be also repaired by B-NHEJ: alternative/backup NHEJ, in which PARP-1 recognizes the DNA ends. The MRN complex and CtIP resect them and finally ligase III together with WRN protein ligate the ends. Modified from Mladenov et al[49].*

2.1.2. Homologous recombination

This pathway is also involved in the repair of DSBs, but in this case using an error-free process. It uses the homologous DNA strand as a template to repair the DNA with high fidelity. Due to the necessity of homologous template, HR is only active during S and G2 phases of the cell cycle, when the sister chromatid is available, although it is theoretically possible that HR also takes place during the other phases of the cell cycle using the homologous chromosome as a template. Nevertheless, the recurrence of HR out of the S-G2 phases is considered a very rare phenomenon due to the compartmentalization of nuclear domains of homologous chromosomes[50]. Another impediment is that it is hypothesized that HR is actively suppressed in G1 cells, in an effort to prevent loss of heterozygosity[51].

Homologous recombination requires three major steps: end resection; strand invasion and finally resolution (Figure 4). Firstly, the MRN complex binds to the DSB where it cooperates with CtIP to promote end-resection. Exonuclease Exo1, DNA2 and BLM helicase are also recruited to the DSB by BRCA1/FANCS where they help in the formation of 3' single-stranded DNA (ssDNA) overhangs. These ssDNA ends are covered by RPA. Then, RPA is replaced by RAD51 in a process that requires multiple mediator proteins such as BRCA2/FANCD1, RAD51B, RAD51C/FANCO, RAD51D and PALB2/FANCN. Once the RAD51/FANCR nucleoprotein filament is formed, it invades the intact double-stranded DNA (dsDNA) from the sister chromatid to search for homology and form a transient structure known as D-loop. When homology is found, RAD51 is released from the 3' overhang of the invading strand and this overhang is used as primer for elongation. Finally, the D-loop can be resolved in two different ways: the most frequent pathway is called synthesis-dependent strand annealing (SDSA) which is characterized by elongation of invading 3'-end throughout a limited distance, followed by displacement of the newly synthesized stretch and re-ligation with the original DNA end, resulting in the repair of the DSB. The other pathway, called double strand break repair (DSBR), is regulated by BLM and needs a second end-capture event, leading to the formation of a double Holliday junction. Depending on the resolution of this junction by BTR complex (formed by BLM, Topoisomerase IIIα, RMI1 and RMI2), it will result in either crossover or non-crossover outcomes[49,52].

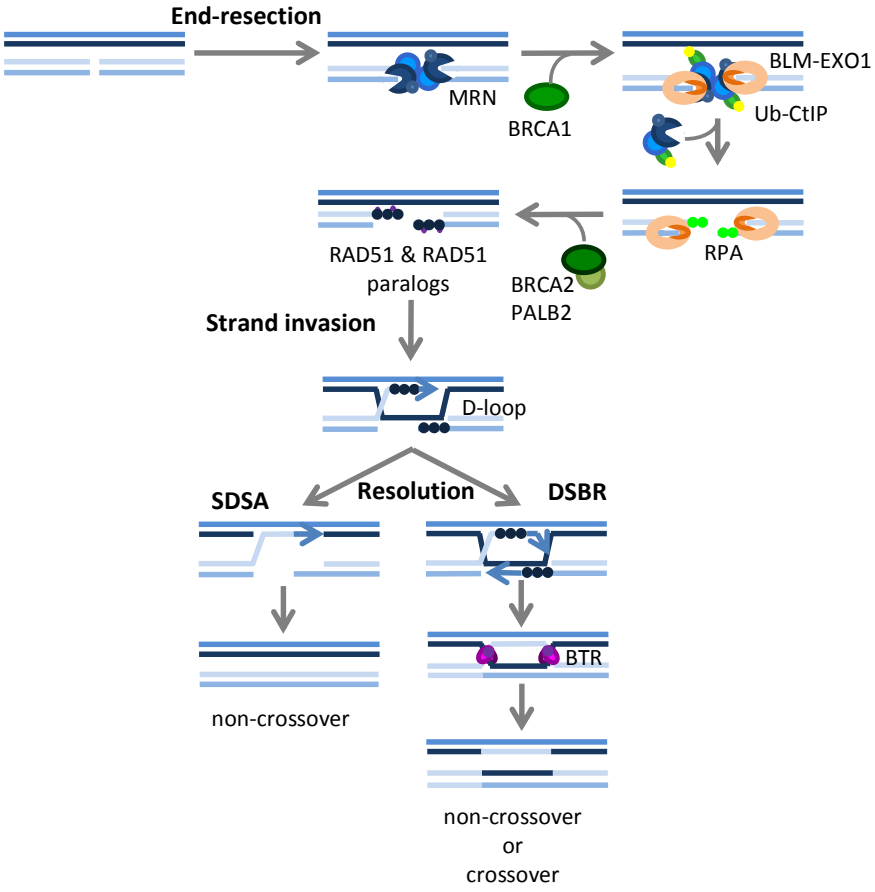


Figure 4: Homologous recombination (HR) pathway *is involved in the repair of double strand breaks when homologous template is available. When a DSB is produced in S-G2 phases of cell cycle the MRN complex binds to DNA ends and BRCA1 protein recruits CtIP which collaborates with MRN in the DNA*

end-resection (BLM-EXO1 are also recruited by BRCA1 and help in the formation of 3' ssDNA overhangs). ssDNA ends are covered by RPA, which is then replaced by RAD51 in a process that requires BCRA2 and PALB2 proteins. RAD51 nucleoprotein filament invades the intact dsDNA from the sister chromatid forming the D-loop when homology is found. D-loop can be resolved by SDSA: synthesis-dependent strand annealing or DSBR: double strand break repair. SDSA is characterized by elongation of invading 3'-end followed by displacement of newly synthesized stretch and re-ligation with the original DNA. DSBR needs a second end-capture event forming a double Holliday junction, which is resolved by BTR complex leading to crossover or non-crossover outcomes. Modified from Mladenov et al[49].

This pathway is strictly regulated, in part by regulation of protein expression at different stages of the cell cycle. Thus RAD51/FANCR, BRCA1/FANCS, BRCA2/FANCD1, BLM and CtIP are present at low levels in G1 and completely absent in G0 cells. The expression of these proteins begins with the start of DNA replication and increases further with the progression of cells through S phase[49]. Another important regulation of HR is done by the phosphorylation of some regulatory proteins after DNA damage. For example CtIP is phosphorylated by ATM, which is a sensor of DNA damage that activates the DSB processing. Another protein that is phosphorylated is BRCA2/FANCD1, this phosphorylation regulates its association with RAD51/FANCR and NBS1, and consequently the DNA damage checkpoints.

2.1.3. Interaction between homologous recombination and non-homologous end joining

The cross-talk between HR and NHEJ is very important, as both pathways repair the same kind of DNA lesion. One example is the regulation of end-resection by 53BP1 and BRCA1/FANCS. 53BP1 inhibits the initiation of DNA end-resection and thus favors repair by NHEJ. On the other hand, BRCA1/FANCS promotes end resection favoring repair by HR. It has been proven that in BCRA1/FANCS deficient cells there is a persistent inhibition of end-resection, which can be abolished by removal of 53BP1[53]. It has also been described a cell-cycle dependent inhibitory function of 53BP1 on BRCA1/FANCS accumulation at DSB in G1 which could also account for suppression of HR in this phase[54]. Finally, there is also a clear link between HR and B-NHEJ since: factors that are important for end resection such as CtIP and MRN complex, mediate HR and B-NHEJ[55,56] and factors that inhibit end-resection such as 53BP1, suppress both pathways[54,57]. Finally, proteins implicated in extensive end-resection such as BLM and Exo1, favor HR over B-NHEJ[55,57].

2.2. Gene targeting mediated by homology-directed repair

Gene targeting mediated by homology directed repair (HDR) needs the introduction of a donor that presents homology to a specific locus and carries the specific sequence to be inserted. The basic structure of this donor consists in two homology arms (HAs) complementary to the region to be integrated, which flank the cassette of interest to be integrated in the specific locus (Figure 5). Although this is the basic structure of a donor, it can also carry selectable markers or promoters depending on the gene targeting strategy. Gene targeting mediated by HDR therefore includes the generation of a DSB in a specific locus by nucleases, the recognition of the complementary regions of the donor and the specific locus, and finally the HDR which facilitates the integration of the transgene in the selected locus of the genome (Figure 5).

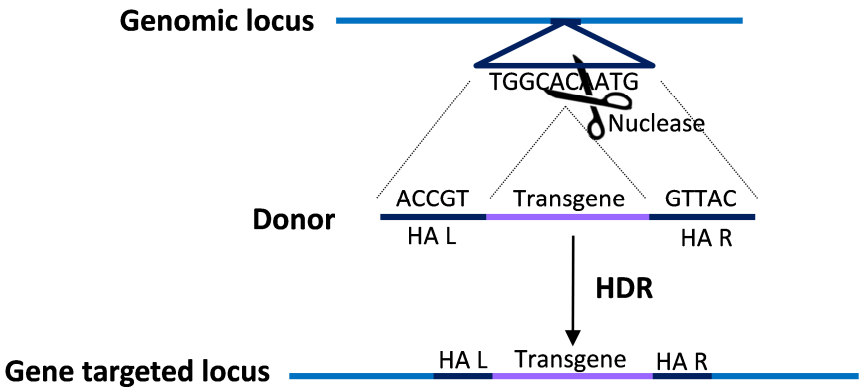


Figure 5: Schematic representation of the donor construct in directed GT before and after integration in a locus. *An arbitrary sequence is represented. The targeted sequence in the genomic locus is cleavage by a nuclease. Since the donor which carries the transgene presents homology to the targeted region in the genome, the endogenous cells machinery of DSB DNA repair can incorporate the transgene in the targeted locus in a process called HDR. HA L: homology arm left. HA R: homology arm right. HDR: homology-directed repair.*

There are mainly three different gene targeting strategies that could be considered to correct a gene by HDR[58] (Figure 6):

a) Gene correction: With this kind of approach one or more bases from the mutated original strand are replaced by integration of the corrected ones. This strategy is the one to be chosen when the goal is to introduce or repair point or small mutations. If the mutations that cause the disease are not recurrent or are produced in different genes, as is the case of FA, this strategy will be restricted to a limited number of patients.
b) Knock-in: A partial cDNA of the gene of interest is introduced in its endogenous locus. Commonly, a splicing acceptor is located before the cDNA to anchor the previous endogenous splicing donor. The endogenous elements of the locus will regulate the expression of the endogenous/exogenous resulting the chimeric cDNA/gene. It is used when it is necessary to maintain the endogenous regulation of transgene[59].
c) Safe harbor integration: Consists in the insertion of a whole expression cassette (promoter, transgene and regulatory signals) in a safe place of the genome. This approach could be used not only for repairing different kind of mutations produced in the same gene, but also for mutations produced in different genes, as it is the case in FA.

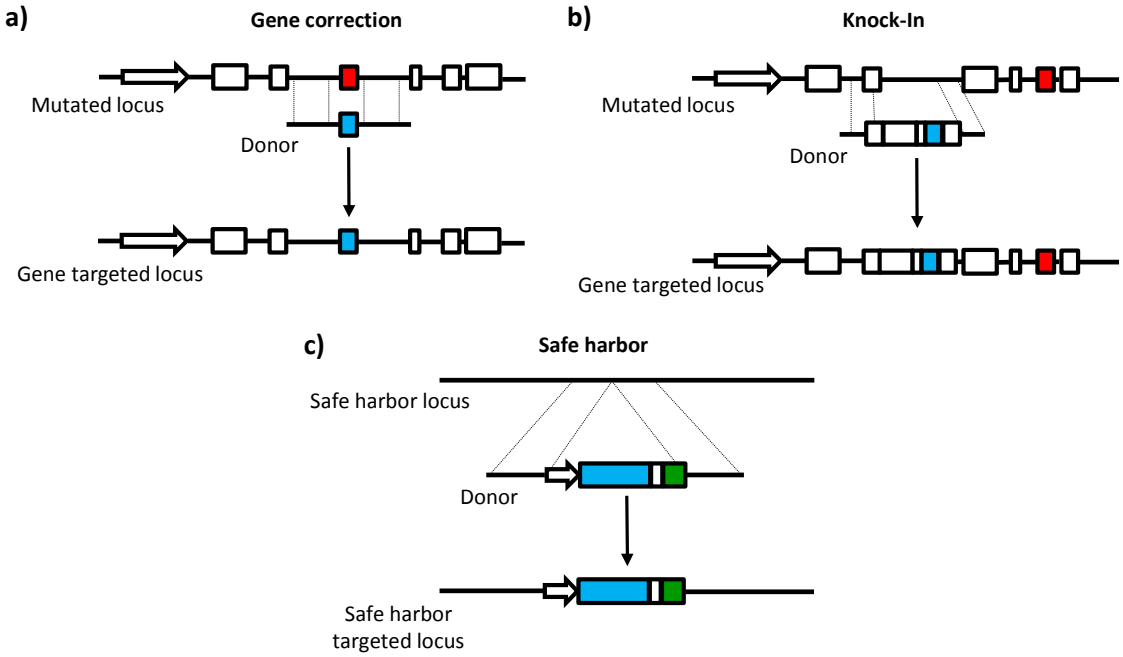


Figure 6: Different strategies for gene editing. *a) Gene correction in which the mutated bases of a specific gene can be replaced. b) Knock-in in which a wild type (WT) part of the cDNA from a specific gene is integrated in a region previous to the mutation. c) Safe harbor in which the transgene with a promoter is inserted in a safe place in the genome. Modified from Garate et al*[58].

2.2.1. Safe harbor loci in the genome

One of the goals of gene targeting strategies is to prevent that the insertion of the transgene in the genome causes insertional oncogenesis or alter the expression of neighboring genes. It is easy to think that the best option should be to integrate the transgene in the locus of the mutated gene (knock-in strategy), as it conserves their endogenous regulation. However, the DSB generated by nucleases in this locus could be in part repaired by NHEJ instead of HDR, and deletions could take place leading to the alteration of the physiological regulation of the gene. Moreover, the insertion of the transgene could lead to alterations in the expression of the neighboring genes and it is not always possible to find specific sequences in the mutated gene for nucleases that cover all the mutations present in a specific disease.

Taking into account all these aspects, an alternative is the insertion of the transgene in a safe harbor place in the genome. It has been demonstrated that the integration of a transgene in these regions do not cause any adverse effect[60]. The main characteristics of safe harbor places are:

- The transgene integrated in this safe harbor can be expressed in all cell types of interest.
- The insertion in this locus or the expression of the transgene should not cause any adverse effect.
- The insertion should not disrupt codifying genes or highly conserved regions of the genome.

- The insertion should not change or alter the expression of the locus and the neighboring loci. Therefore, the interaction between enhancers and promoters in the transgene and the locus and neighboring areas should be as minimal as possible.

Safe harbor sites could be allocated in genes or in genetic deserts. Both strategies present advantages and disadvantages. The main advantage of considering a locus as a safe harbor is that the chromatin conformation in this area is usually opened, allowing the expression of the integrated transgene. However, the transgene incorporation could trans-activate the endogenous genes or deregulate its expression due to the disruption of a regulator area in the intron in which they are usually incorporated. Moreover, genes that could be used as safe harbor locus should be expressed widely all over cell types or at least in the cells of interest. On the other hand, genetic deserts do not present in principle, the problem of transactivation or deregulation of endogenous genes, although these areas could contain regulatory regions driving expression of very distal genes. The main limitation of these areas is that the chromatin is usually very compacted, making gene targeting more difficult and also reducing the expression of the integrated transgenes. Moreover, potential conformational changes produced in the chromatin in these areas after gene editing should be taken into account in terms of a possible deregulation of the cell status.

Safe harbor sites should be functionally validated, enhancing the effect of the insertion in neighboring genes *in vitro*, and determining possible changes in the epigenetic marks in the locus after the introduction of the transgene. Finally, these *in vitro* studies should be confirmed *in vivo* to conclude that the targeting in this locus does not produce any alteration in the tissue and in the organism, including risks of malignant transformation.

The two safe harbor most frequently used are the chemokine (C-C motif) receptor 5 (*CCR5)* and the adeno-associated virus site 1 (*AAVS1)*:

a) *CCR5*: is located in chromosome 3 (3p21.31) and codifies the principal co-receptor of the HIV-1. The first evidence that pointed out that this locus could act as a safe harbor came from the resistance of individuals to HIV-1 infection, due to a null mutation in this gene (*CCR5Δ32*)[61]. The mutation of this gene did not induce any pathology and has been already used in the first clinical trial using nucleases to treat HIV-1 patients; in which no side effects have been observed so far[62]. The endogenous expression of this gene is variable along cell types: high in T lymphocytes, monocytes and macrophages, but it is not expressed in B lymphocytes or dendritic cells[63]. Moreover, the knockout mice for *Ccr5* presents impaired leukocyte migration and increased susceptibility to some infections[64]. Another disadvantage of this locus is that the expression of a reporter gene, *GFP*, in T cells is lower in this site as compared with the expression of the same reporter gene in another safe harbor locus: the *AAVS1*[65]. Furthermore, the high homology of this locus with the *CCR2* gene make the design of nucleases more complicated increasing the off-target events[66].
b) *AAVS1*: is located in chromosome 19 (19q13.42). It was described as the common integration site of the adeno-associated human virus[67], which disrupts the protein phosphatase I regulatory inhibitor subunit 12C (*PPP1R12C*). This gene codifies a protein whose function is not well understood, and no side effects have been

described derived from its interruption. Moreover, it has been described that approximately 80% of the population presents detectable levels of antibodies against some AAV serotypes, which means that they probably harbor integrations of this virus in the *AAVS1* locus without any pathogenic or side effect[68]. This region has been used in gene targeting experiments in T-cells, neural stem cells, embryonic stem cells, fibroblasts and iPSC[69-75]. These studies support a stable and robust expression of the different transgenes integrated in the *AAVS1* locus all over the different cell types. Moreover, it has been demonstrated that the integration of a transgene in this locus does not induce any adverse effect[65]. The chromatin of this locus has an open and active conformation, and the insertion in the *AAVS1* locus does not alter the expression of surrounding genes, probably due to the presence of an insulator, which prevents the transactivated expression of surrounding genes[76]. All these characteristics make the *AAVS1* locus as the best safe harbor locus for gene targeting approaches described up to now.

2.3. Designed nucleases

Nucleases are the tools that generate the DSBs in a specific locus in the genome for gene targeting. There are four different types of nucleases that have been widely used in gene targeting: meganucleases (MGNs), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered, regularly interspaced, short palindromic repeat associated to Cas9 nuclease (CRISPR/Cas9) system (Figure 7). All of them possess a DNA recognition region and a catalytic domain which generates the DSB in a specific locus.

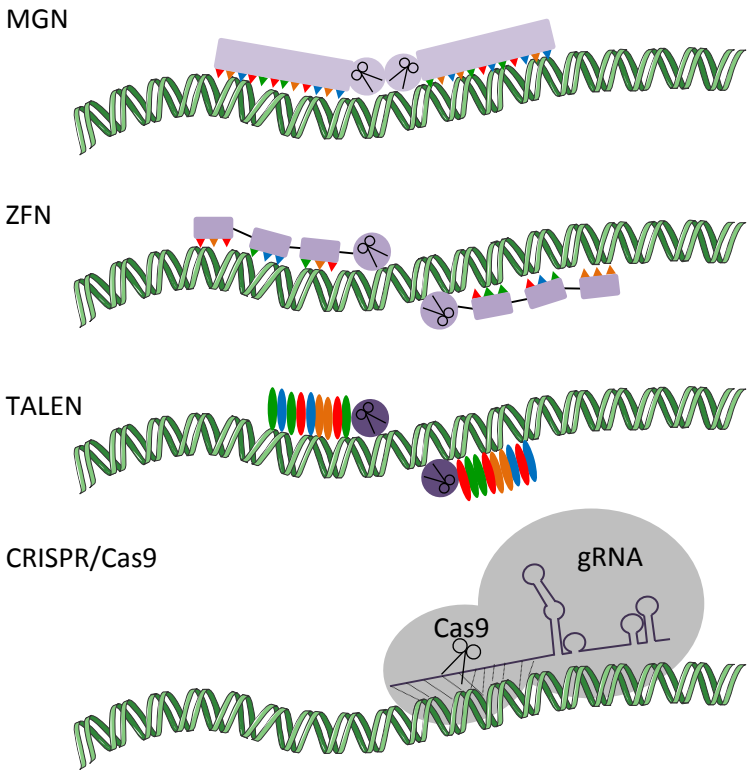


Figure 7: Different types of nucleases. *MGN: meganuclease. They recognize long sequences of DNA cutting by homology. ZFN: zinc-finger nuclease. They are formed by two subunits: homology domain composed by groups of three ZF domains which recognize a specific nucleotide and the catalytic domain*

formed by FokI endonuclease which needs to dimerize in order to cut. TALEN: transcription activator-like effector nuclease. They are also formed by a catalytic domain composed by the FokI endonuclease and a DNA binding domain formed by a tandem of highly repeated sequences which recognizes each nucleotide. CRISPR/Cas9 systems are formed by Cas9 protein which presents the cleavage activity and a gRNA which recognizes a specific region in the genome and directs the cut of the Cas9.

Although they are designed to recognize a specific sequence in the genome, they could also produce DSBs in other parts different from their target sites that have been called off-targets. Although this phenomenon is rare, it could decrease the cutting efficiency of the nucleases in the on-target sites and increase the genomic instability of the cells[77]. This off-target activity of the nucleases is linked to regions in the genome that share a high homology to the target sites. This issue needs to be studied deeply when a new nuclease is designed. Due to the deep characterization[65] and the previous experience that our laboratory had with ZFNs designed by Sangamo Biosciences (Richmond, California, USA) to target the *AAVS1* locus[75], our current studies were performed with this nuclease.

2.3.1. Meganucleases

MGNs belong to a family of natural endonucleases that are present in all organisms. They recognize long sequences of DNA (between 12-45 bp)[78,79] cutting the homology arm that do not contain the recognized sequence[80]. The main limitation of these enzymes is that they can only be designed to cut regions in which the restriction site of the enzyme is present. Despite the identification of hundreds of different wild type (WT) MGNs, their target sites do not cover the entire human genome. To overcome this problem engineered MGNs are produced by modifications of the WT ones, fusing different domains to generate chimeric proteins[81]. However, and despite the description of some artificial MGNs, their design is very complicated and their use is restricted. There is only one paper in which these enzymes are used to target murine HSCs[82].

2.3.2. Transcription activator-like effector nucleases

TALENs were firstly described in a *Xhantomonas* pathogen[83]. They are formed by a catalytic domain composed by non-specific DNA cleavage domain of the *Fok*I restriction enzyme joined to a specific DNA binding domain[84]. The DNA binding domain of these proteins is formed by a tandem of 10-30 highly repeated sequences. The specificity of each repeat unit is driven by polymorphisms at positions 12 and 13 called repeat-variable di-residues (RVDs) which confers specificity of each repetition to a single nucleotide on the target sequence[85]. The recognition of the target sequence is therefore based on the affinity of the amino acids (aa_s) sequence and the DNA sequence. A lot of different TALENs have been generated since the code of TALE-to-DNA interactions was discovered in 2009[86] due to their relatively easy design[84]. However, the highly repeated nature of these enzymes hampers their packaging into viral vectors for delivery and their size (more than 3 kb) complicates their delivery into cells.

The catalytic part of the enzyme is given by the *Fok*I enzyme. Since this restriction enzyme is only active as a dimer[87], these artificial nucleases are composed of active pairs in which the two monomers bind adjacent target subsites separated by a DNA spacer that allows for the formation of an active dimer to cleave the target locus[88].

Although not many studies have compared TALENs and ZFNs directly, two studies suggested that TALENs and ZFNs present similar gene targeting efficiencies but TALENs produce less off-target mutagenesis than ZFNs[66,89].

2.3.3. CRISPR/Cas9 system

The CRISPR/Cas9 system is part of the adaptive immune system in prokaryotes to protect these cells against viral infections or plasmids[90,91]. This system incorporates sequences from invading DNA between CRISPR repeat sequences encoded as arrays within the bacterial host genome. Transcripts from the CRISPR repeat arrays are processed into CRISPR RNAs (crRNAs), each harboring a variable sequence transcribed from the invading DNA, known as the "protospacer" sequence, and part of the CRISPR repeat. crRNAs are then used to target and destroy any foreign DNA sequences that enter the cells. The bacterial Cas9 nuclease and a crRNA form a ternary complex with a second RNA component, the trans-activating crRNA (tracrRNA), which has a fixed sequence. The protospacer-encoded portion of the crRNA directs Cas9 to cleave complementary target sequences. If they are adjacent to short sequences known as protospacer adjacent motifs (PAMs), Cas9 introduces a DSB 3 bp upstream of the PAM[92]. The RNA components of the CRISPR/Cas9 system can be combined into a single guide RNA (gRNA)[93] that is delivered in a simple two-component system with the Cas9 in cells to conduct gene targeting approaches. The advantages of the CRISPR/Cas9 system are the high efficiency to generate DSBs into the genome of mammalian cells and the simplicity of its engineering. The Cas9 nuclease is used to target different regions of the genome and only the protospacer region (20 nucleotides) of the gRNA needs to be modified. The off-target activity of CRISPR/Cas9 system seems to be higher than the other nucleases[94], and a lot of efforts are nowadays being done to unravel this problem. One example is the reduction of gRNA protospacer length up to three nucleotides, which appears to make the gRNA less tolerant to mismatches and less likely to bind to the off-target sites[95].

2.3.4. Zinc-finger nucleases

Zinc-finger nucleases are artificial enzymes that were formed by engineered Cys_2-Hys_2 zinc finger (ZF) domain fused to the nuclease domain of the restriction enzyme *Fok*I. In this configuration, the DNA-binding ZF domain directs the nonspecific *Fok*I cleavage domain to a specific DNA target site. Even though academic consortia have developed open-source libraries for ZFN design[96,97], engineering of site-specific ZFNs remains difficult for non-specialists.

The ZF domain is the most common DNA binding motif in eukaryotes. It consists in 30 residues that fold into a ββα-structure coordinated by a zinc ion[98]. Each ZFN is composed by 3-4 ZF domains organized in tandem arrays which recognize a target DNA sequence from 9 to 12 bp; and at the same time each ZF domain presents an α-helix which recognizes a specific DNA triplet. As the cleavage activity of *Fok*I requires dimerization and it has been demonstrated that this dimer cut more efficiently when enzymes are in inverted orientation[88], ZFNs are designed to work as pairs of two monomers in reverse orientation. Each monomer binds to half-sites that are separated by a spacer sequence. Due to the off-target activity observed in some studies using ZFNs[99,100], variants of the endonucleases have been developed to overcome this problem. These consist of a mix of two distinct ZFNs with different *Fok*I domains that are obligated heterodimers, so each ZFN pair introduces a DSB only when the two distinct ZFNs are able to bind adjacent DNA regions[101,102].

The origin of these nucleases was in 1996[103]. From the very beginning ZFNs have been used in a high variety of species for different applications:

a) Genetic inactivation or interruption: this process occurs when the repair of the DSB generated by ZFNs is mediated by NHEJ. It has been used to silence a target gene. This tool has been extensively used to generate knockout models in different species: zebrafish[104], mice[105], rats[106,107], rabbits[108] or pigs[109]. However, the most important work in this direction is related to the treatment of the human immunodeficient virus type 1 (HIV-1) infection. These studies are based on the disruption of *CCR5*, one of the co-receptors used by the HIV to enter into the cells. Preclinical studies have been conducted in $CD4^+$ T lymphocytes, which are the main target cells for the virus[110] or in HSPCs[111]. Based on the good results obtained in these preclinical studies, seven different clinical trials have been or are been performed based on this strategy. One of them, a phase I clinical trial (NCT00842634) has been already performed using autologous genetically modified $CD4^+$ T-cells that were injected in HIV-1 infected patients. This phase I clinical trial has shown the safety of this approach in patients[62].
b) Safe harbor: ZFNs have been also used to insert a gene in a specific site of the genome, with the aim of evaluating the safety of the insertion and correcting *in vitro* the phenotype of specific diseases. For example the *GFP* reporter gene has been inserted in the *CCR5*[112] and the *AAVS1*[65,113] in cell lines[65,112] and primary cells[113]. As it has been mentioned, the insertion of therapeutic genes in safe harbor loci has also been conducted in different cell types such as primary fibroblasts[75], induced pluripotent stem cells (iPSC)[114] or $CD34^+$[113] (antigen that identifies hematopoietic stem and progenitor cells (HSPCs), which marks a heterogeneous cell population with both short- and long-term repopulating cells). ZFNs designed for the *AAVS1* locus are our selected tools to conduct our gene targeting experiments as they have been deeply characterized and are very specific[69,99]. However, although in previous reports no off-targets using this ZFNs have been reported, one study in which the ZFNs designed by Sangamo Biosciences were compared with TALENs targeting the same region, the authors claim that *AAVS1* ZFNs are less specific than TALENs[66].
c) Gene correction: in this case the correction takes place in the mutated locus producing the substitution of the particular mutation, for this reason it requires the design of specific ZFNs that needs a deep study of their efficiency and specificity. This approach has been used in HSPCs from patients with in sickle cell disease, to correct a single point mutation in *β-globin* gene[115].
d) Knock-in: the correction, as in the case of the previous strategy, takes place in the mutated locus, but this time introducing a partial cDNA of the mutated gene instead the correction of a specific mutation. This was the strategy selected to correct HSPCs from X-SCID patients in order to correct mutations in the interleukin 2 receptor gamma (*IL2RG*) gene[113].

2.4. Nuclease and donor delivery methods

The delivery of nucleases and donors is nowadays one of the key challenges in gene targeting[116]. The main objective of delivering the nucleases effectively in the cells is to promote the activity of the enzyme to allow the generation of DSBs efficiently in a very short period of

time to avoid toxicity and off target activity due to the long term expression of the nuclease. Nowadays, the preferred approach is based on the delivery of donor as plasmid DNA, and the nuclease as plasmid DNA or *in vitro*-transcribed mRNA. Integrase-defective viral vectors also serve as a robust source to facilitates the delivery of either the nuclease or the donor construct in cell types that are very sensitive to transfection such as HSPCs[117].

The delivery of plasmid DNA is widely used[118,119], as this molecule is easily engineered and its transfection in cultured cells is a standard technique carried out in laboratories all over the world. Linear donor sequences with less than 50 bp of homology[120], as well as single stranded DNA oligonucleotides[121] can also be used to induce mutations or correction at the target site. However the use of this kind of DNA structures limits the transgene size. Due to the negative charge of both plasmid DNA and the cell surface, the uptake of DNA into the cells is restricted, and the delivery needs to be supported by complexing DNA with chemicals such as calcium phosphate, coat lipids or cationic polymers[118], or by subjecting the cells to an electric field which open pores in the nuclear and/or plasmatic membrane of the cells (electroporation). Nucleofection - a kind of electroporation - is a standard method for transferring DNA to hard-to-transfect cell types such as HSPCs. However, this process could produce high cellular toxicity and even cell death. Moreover, the delivery of plasmid DNA has a risk of mediating random insertion into the genome of treated cells[122]. The risk of random plasmid integration depends on the plasmid quality, but it is also influenced by properties of the targeted cells, including the prevalence of DSBs and the status of their DNA repair pathways. Plasmid integration, although is a non-usual event, represents an obstacle for gene targeting approaches.

Considering the risk of inserting plasmid DNA into the genome of treated cells, the transfection of *in vitro*-transcribed mRNA has been used as an alternative delivery of nucleases[123,124]. The main advantage of this kind of delivery is that mRNA establishes a short-term boost of enzyme activity with no risk of integration. Moreover, transfection of mRNA in cells is a less toxic process than DNA, as foreign DNAs generate more acute cellular responses than foreign mRNAs probably due to their higher stability in the cell. However, mRNA, as happens with DNA, is a biomolecule negatively charged and has the same requirements to enter in a cell as DNA. Another possibility is the delivery of nucleases as proteins; this approach presents the same advantages as mRNA delivery, since they also act in a hit-and-run fashion. Genetic fusion of recombinant proteins to a positively supercharged moieties called protein transduction domains (PTDs), favors their uptake by cellular internalization mechanisms[125,126]. However, the generation of high yields of soluble and active PTD-containing nucleases is difficult[127]. An alternative approach consists in the chemical conjugation of nucleases to PTDs for receptor-mediated endocytosis[128]. Interestingly, owing to the net positive charge of their Cys_2-His_2 zinc-finger motifs, ZFNs display an intrinsic cell penetrating capacity, which can lead to targeted mutagenesis in a variety of cell types[127]. Other delivery options under investigation include protein transfection procedures such as electroporation[129]. Chemical transfection agents are also being investigated for nuclease as protein direct transfer[130].

Viral delivery has been used as a key instrument in conventional genetic therapies[4] and hold a great potential also for delivery of genome-modifying tools. Viral vectors not only constitute efficient gene transfer tools, but can also be used as protein and mRNA transfer tools[131,132]. Four different types of viral vectors have been used to deliver both nucleases and donors in

gene targeting experiments: baculoviral vectors (BVs), adeno-associated vectors (AAVs), adenoviral vectors (AdVs) and integrase-defective lentiviral vectors (IDLVs). BVs are known to infect invertebrates, but have been also tested in gene editing experiments involving the delivery of donor DNA and ZFNs in embryonic stem cells[133]. AAVs naturally persist in human cells and presents nonpathogenic character. The major problem of AAVs is the restriction in their packaging capacity that is incompatible with some therapeutic gene sizes. However some AAVs have been adapted, avoiding its integrative nature, for nuclease and donor delivery in human cell lines, embryonic stem cells and hemophilia B *in vivo* murine models[134-137]. AdVs and IDLVs do not integrate in the genome of the targeted cell. AdVs are capable of packaging up to 37 kilobases (kbs) of DNA, however the transduction capacity of these vectors is highly dependent on the cellular type. They have been used in some gene targeting approaches to deliver the nuclease both in fibroblasts and $CD4^+$ T cells[75,110]. Finally, IDLVs allow the expression of the donor in the cells due to the DNA intermediates that are generated after transduction both in cell lines and primary cells[65,75,112,113].

2.5. Gene targeting improvement

Although with the emergence of nucleases the efficacy of gene targeting has improved, a big effort is needed in order to make this kind of approach more efficient for its use in the clinic. As we have previously mentioned, the delivery of the nucleases and the donor is a bottle neck for gene editing of hematopoietic stem cells, because these cells are hard to transfect and the only methods available so far to deliver nucleases and donor constructs are nucleofection, which is very toxic; and transduction with viral vectors. Other point that should be considered is that although the generation of a DSB in the genome with nucleases and the introduction of a donor favor the repair of the DSBs by HDR, this DNA repair pathway is still less frequent than NHEJ. Different studies have been developed in order to overcome these problems and to increase the efficiency of gene targeting, some of which have been studied during this thesis.

2.5.1. Improvement of the transduction efficacy: nanoparticles

One of the keys for the successful HSC GT is the development of methods that allow a highly efficient gene delivery under *ex vivo* conditions that do not significantly alter the biological properties of the HSCs. It is known that the short-term culture of $CD34^+$ cells with the appropriate cytokines approximately doubles the transduction efficiency as compared to efficacies achieved under non-stimulating conditions[138]. However, the prolonged culture with cytokines results in a loss of human cell engraftment in immunodeficient mice, suggesting the necessity of achieving a compromise between the transduction efficiency and the engraftment potential during *ex vivo* culture. Some methods that have been widely used to enhance the LV transduction in HSCs are based on the increase of the multiplicity of infection (MOI) or the use of sequential transductions, but this cannot always be accomplished, mainly due to the limited quantity of LVs. One mechanism that has been recently shown to increase the transduction efficiency of lentiviral vectors is the brief *ex vivo* treatment of transduced cells with rapamycin; an inhibitor of mTOR complex[139,140]. It has been proven that this compound increases the LV entry, leading to increase levels of reverse-transcription and genomic integration. This increase in transduction also correlates with an increment in the transduction efficiency of the most primitive compartment of HSCs, preserving the long-term engraftment of these cells. This increment in transduction also been observed when IDLVs are used[140]. These studies pointed

out that the cytoplasmic entry of the LV could be the biggest restriction for the transduction of HSCs with LVs and rapamycin can solve the problem.

Another possibility could be focus in the use of nanoparticles (NPs). They are small particles (between 1 and 100 nm in size) that could be formed by different inorganic materials. For biological applications of NPs, it is necessary their functionalization with one or several biomolecules, such as DNA/RNA, oligonucleotides, peptides and antibodies, fluorescent dyes, polymers, drugs, tumor markers, enzymes and other proteins that will introduce the required bio-functionalities[141]. All NPs are normally functionalized at least with polyethylene glycol (PEG) as this molecule stabilizes the NP and prolongs their half-life (Figure 8). NPs act as multifunctional devices capable of bypassing biological barriers for specific delivery of therapeutic agents. Among the numerous inorganic nanoparticles that are available nowadays, gold nanoparticles (AuNPs) are optimal non-toxic carriers for gene therapy due to their chemical properties and ease of surface modification with a variety of ligands[142]. Different studies using engineered AuNPs modified with small interfering RNA (siRNA) have demonstrated *in vitro* and *in vivo* cytoplasmic delivery of siRNA and efficient gene silencing[143,144]. Conde and co-workers[145] reported functionalization of siRNA by ionic coupling to a positively charged layer formed by quaternary ammonium groups (R_4N^+). Ionic interactions between the negatively charged siRNA backbone (via phosphate groups) and quaternary ammonium positively charged groups ensured binding of siRNA onto the AuNP's surface. This approach could be used to deliver any molecule that is negatively charged, as the viral particles. The ionic binding rate mainly depends on the amount of charges present on the NPs and the biomolecules. Moreover, there are some studies in which the combination of AdV with gold/iron-oxide magnetic nanoparticles increase the transduction efficiency of the viral vectors and allow gene transfer to adeno-resistant cells by using a magnetic field[146].

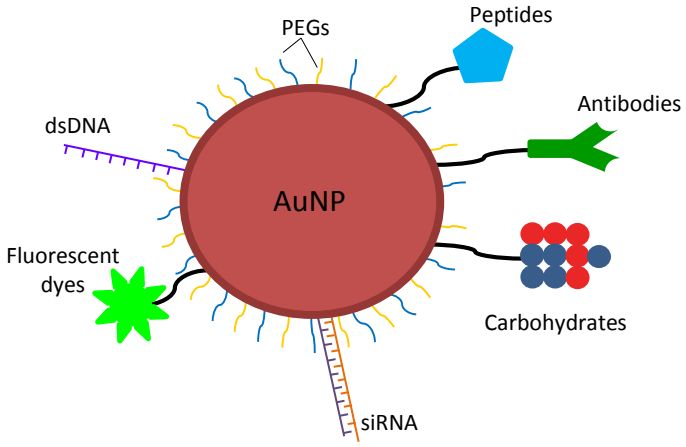


Figure 8: Schematic representation of gold nanoparticles (AuNP). *a) Multifunctional AuNP which is functionalized with polyethylene glycol (PEG), peptides, antibodies, carbohydrates, siRNAs, fluorescence dyes or dsDNA. Modified from Conde et al[141].*

In the present thesis we have proposed to study the possible application of these AuNPs to enhance the transduction efficiency of IDLVs and therefore the gene targeting in general.

2.5.2. Improvement in homology-directed repair frequency: PARI inhibition

As previously mentioned, when a DSB is produced by natural or artificial nucleases, NHEJ is the preferential via to repair this DNA damage. Therefore, improving the efficiency of precise gene targeting remains a major challenge. A recent study reported the function of small molecules for enhancing the efficacy of gene targeting[147]. These compounds inhibit NHEJ or potentiate HDR when added to the medium after transduction with both CRISPR/Cas9 and a reporter donor in different cell types. Moreover other groups has inhibited key components of NHEJ such as DNA ligase IV, during their gene targeting experiments in order to increase the efficiency of HDR[148,149]

Consistent with this idea, there is a previous study in which the transient downregulation of a PCNA-interacting factor, called PCNA-associated recombination inhibitor (PARI) improves homologous recombination and genomic stability in BRCA2/FANCD1 deficient cells, where HR is impaired[150]. PARI is an anti-recombinase which suppresses inappropriate HR. This protein is recruited in S phase of the cell cycle in a PCNA-dependent manner to inhibit recombination events at replication forks. In healthy donor (HD) cells PARI removes RAD51 from DNA when is recruited preventing hyper-recombination events that leads to genomic instability (Figure 9). However, in cells with mutations in some HR critical factors such as BRCA2/FANCD1 deficient cells, in which the frequency of HR events is very low, the inhibition of this anti-recombinant factor increases HR to normal levels and results in protective effects, suppressing genomic instability.

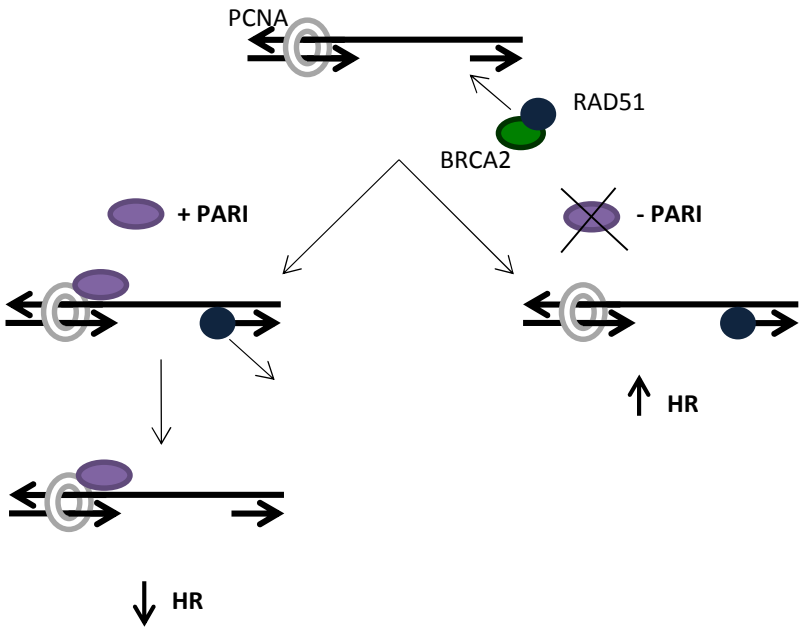


Figure 9: PARI model in cells from healthy donors. *In S phase of the cell cycle RAD51 is recruited to the sites in which the DNA is been replicating. If PARI protein is present (left), it binds to PCNA and mediates the release of RAD51 from the replicating DNA controlling the level of HR. However, if PARI protein is not present (right), RAD51 is still bound to the DNA ends of replicating strands and the frequency of HR is increased.*

Since the inhibition of PARI in BRCA2 deficient cells increases the efficiency of HR, one of the objectives of this thesis will be to study the role of PARI inhibition in gene targeting efficiency in cells from FA-A patients where HR is mildly affected in comparison with BRCA2 deficient cells[151].

3. Fanconi Anemia

3.1. General characteristics of the disease

Fanconi anemia (FA) is a genetic inherited disorder that was first described in 1927 by the Swiss pediatrician Guido Fanconi in a family with five children, three of them presenting physical abnormalities and hematological defects[152]. Posterior studies performed in the peripheral blood (PB) of these and other patients with similar characteristics allowed him to appreciate that the disorder affects all the hematopoietic lineages, being these alterations the main cause of death[153].

Fanconi anemia is a highly heterogeneous syndrome mainly characterized by congenital abnormalities, hematological defects and increased cancer predisposition. Among these symptoms, bone marrow failure (BMF) is the most frequent cause of death in FA pediatric patients (50% of the patients develop BMF during the first decade of their lives)[154,155].

The disease is caused by mutations in any of the 19 genes described so far (Table 2). Depending on the mutated gene, patients are assigned to different complementation groups (FA-A to FA-T). Fanconi anemia is an autosomal-recessive disorder with the exception of patients with mutations in *FANCB* which is X-linked. All these genes are members of a DNA repair pathway known as the FA/BRCA pathway, involved in the resolution of DNA interstrand-crosslinks (ICLs)[156-158]. Mutations in any of these genes increase the genetic instability of cells and compromise their survival upon exposure to ICL agents. These compounds include metabolic products such as aldehydes, which come from alcohol detoxification or lipid peroxidation; or exogenous agents such as cysplatin, mitomycin C (MMC) or diepoxybutane (DEB)[159,160]. The hypersensitivity of FA cells to ICL agents has been used as a diagnostic marker for the disease[161].

Although mutatins in 19 genes have been described to cause the disease, only 15 of them (*FANCA*, *B*, *C*, *D1*, *D2*, *E*, *F*, *G*, *I*, *J*, *L*, *N*, *P*, *Q* and *T*) fit all the criteria to be considered bona fide FA genes: mutations in these genes have been found in at least two patients who present BMF, and their cells are characterized by chromosomal instability measured by a positive chromosome fragility test. Mutations in *FANCO*, *FANCR* and *FANCS* genes results in a chromosome fragility syndrome without BMF, and are considered FA-like genes. Moreover, some recent publications claim that *FANCM* should not be considered a FA gene[154] since there is only one patient reported with mutations in this gene that interestingly also harbored biallelic mutations in *FANCA*[162]. Additionally, the ectopic expression of FANCM in this patient did not rescue the FA phenotype. Finally, it has been described that two individuals with loss of function of *FANCM* are healthy and exhibit normal hematology[162,163]. The FA/BRCA pathway also involves other associated proteins: FAAP10, FAAP16, FAAP20, FAAP24, FAAP100 and

FAN1; that although cooperate in the FA pathway, their mutations have not been found in any FA patients and therefore are not considered FA proteins[157].

FA gene	Prevalence	Chromosomal location	Protein size (kDa)	Protein characteristics
FANCA[164,165]	60-70%	16q24.3	163	Member of the core complex
FANCB[166]	2%	Xp22.2	95	Member of the core complex
FANCC[167]	14%	9q22.3	63	Member of the core complex
FANCD1/BRCA2[168,169]	3%	13q12-13	380	Involved in HR. Cancer susceptibility gene
FANCD2[170]	3%	3p25.3	155	Member of the FA ID2 complex. Monoubiquitinated by FANCL (core complex). Directs downstream events
FANCE[171]	3%	6p21.3	60	Member of the core complex
FANCF[172]	2%	11p15	42	Member of the core complex
FANCG/XRCC9[173,174]	10%	9p13	68	Member of the core complex
FANCI[175-179]	1%	15q26.1	149	Member of the FA ID2 complex. Monoubiquitinated by FANCL (core complex) in a FANCD2 dependent manner. Directs downstream events
FANCJ/ BRIP1/ BACH1[24,29-31]	2%	17q23.1	140	Involved in HR: 5'-3' DNA helicase. Cancer susceptibility gene
FANCL[180]	0.2%	2p16.1	43	FA core complex. E3 ubiquitin ligase
FANCM[181]	0.2%	14q21.3	250	Member of the core complex. DNA translocase. Important for ATR activation during ICL repair. Cancer susceptibility gene
FANCN/PALB2[182-185]	0.7%	16p12.1	140	Involved in HR. Cancer susceptibility gene
FANCO/RAD51C[186]	0.2%	17q22	43	Involved in HR. Cancer susceptibility gene
FANCP/ SLX4/ BTBD12[187,188]	0.2%	16p13.3	200	Scaffold/regulator of XPF-ERCC1, MUS81-EME1 and SLX1 nucleases
FANCQ/ERCC4/ XPF[189]	NR	16p13.12	104	Associates with ERCC1 to form a FANCP-dependent ICL unhooking nuclease. Participates in NER independently of FANCP
FANCR/RAD51[190]	NR	15q15.1	37	Involved in HR. Interacts with FANCS and FANCD1
FANCS/ BRCA1[191]	NR	17q21	208	Involved in HR. Inhibition of NHEJ. Removal of CMG (CDC45-MCM-GINS) complex during ICL repair. Cancer susceptibility gene
FANCT/UBE2T[192,193]	NR	1q32.1	22.5	FA core complex. E2 ubiquitin conjugating enzyme

Table 2: FA genes described up to now. *The table shows the prevalence in USA (NIH database)[194] of the mutated genes (last review in 2013), chromosomal location, protein size and the main characteristics of each protein. NER: nucleotide excision repair. HR: homologous recombination. NHEJ: non-homologous end-joining. ICL: interstrand-cross link. NR: not reported.*

3.2. Fanconi Anemia/BRCA pathway and the repair of interstrand-crosslink lesions

Interstrand-crosslink lesions impede the normal progression of DNA replication or transcription because the DNA strands cannot be separated owing to the covalent linkage between the two strands. These processes are essential for correct cell function, so ICLs should be repaired during all phases of cell cycle. The FA/BRCA pathway is reported to be active following the stalled replication fork generated by an ICL or during S phase entry[195]. In the absence of active replication forks and a homologous template, ICL repair depends on nucleotide excision repair (NER), especially transcription-coupled NER (TC-NER) and translesion-synthesis (TLS) proteins[196,197]. Other DNA repair factors are also activated in response to this type of DNA damage, such as ATR, ATM or the effector kinase CHEK1. These proteins arrest the cell cycle[198] in order to repair the DNA damage.

The first protein that initiates the DNA repair process is FANCM. This protein presents translocase activity[181] and, together with FAAP24 and MHF1/2[199], moves across the DNA to detect stalled replication forks[200], facilitating the translocation and anchoring of the rest of the FA core complex proteins (FANCA, B, C, E, F, G, L and T) to the damaged DNA. Moreover, FANCM also activates ATR, which phosphorylates some members of the core complex like FANCE and FANCG leading to the activation of FA/BCRA pathway[201].

Once the core complex is recruited to the ICL-induced stalled replication fork, FANCL, the catalytic E3 ligase subunit of the core complex, in concert with FANCT/UBE2T[192], ubiquitinates the FANCD2-FANCI (ID2) complex. The mono-ubiquitination takes place on lysine 561 of the FANCD2 protein and on lysine 523 of the FANCI protein[177,202]. This process produces the heterodimerization of these two proteins. Mutations in any of the core proteins of the pathway - with the exception of FANCM, which present a redundant activity in other proteins such as the mismatch repair complex MutS[203] - lead to a deficient mono-ubiquitination of FANCD2 and FANCI.

The ubiquitinated ID2 complex is translocated to the DNA repair sites, forming nuclear repair foci with other downstream proteins. Once in the DNA repair foci it recruits FANCQ-ERCC1-FANCP-SLX1 - where FANCP serves as a master scaffold - and 5' and 3' incisions on one parental strand are generated, allowing unhooking of DNA where the ICL has occurred. Other nucleases, including the FAN1 and SNM1 family nucleases might act redundantly to make incisions at the ICL[204]. Unhooking converts the stalled replication fork in a double strand break (DSB) and allows translesion polymerases to repair one strand. Polymerase ζ and Rev1 are the key proteins in this step of ICL repair[205,206]. Rev1 acts as a scaffold, facilitates polymerase exchange[207] and inserts a dCMP residue opposite to the ICL[208]; Pol ζ extends the DNA from the distorted primer-template termini (those formed by an insertion of a nucleotide at a lesion by another TLS polymerase). TLS polymerases need the monoubiquitination of PCNA - a polymerase processivity factor that encircles DNA and functions as a moving platform for DNA synthesis[209] - to be recruited to the lesion[210].

The repaired strand can then be used for repairing the other strand using homologous recombination (HR). The activation of the ID2 complex also moves FANCD1 to chromatin, interacting with FANCJ, FANCN and FANCS, and thus facilitating the assembly of DNA damage-

inducible FANCR nuclear foci. All these remaining FA proteins: FANCD1, J, N, O and S together with FANCR, are implicated in the resolution of DSBs by HR (see section 2.1.2. of the introduction).

Finally the unhooked ICL, which no longer represents an obstacle for the completion of S-phase, is removed by NER[159] and the deubiquitinating enzyme, USP1 deubiquitinates the ID2 complex to finish the process. Defects in this event have also been associated with ICL sensitivity[211].

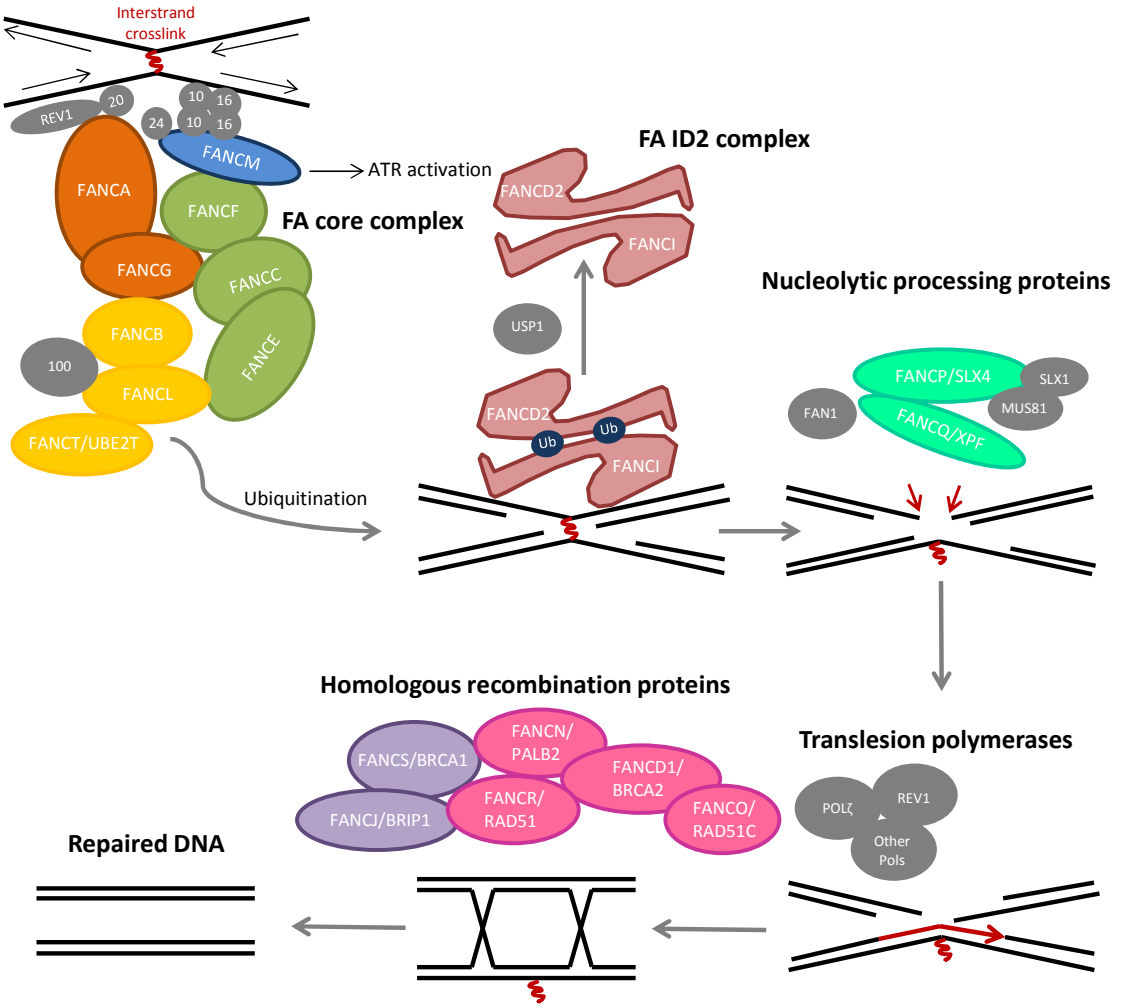


Figure 10: FA/BRCA pathway. *Proteins involved in the different steps of ICL repair are shown. In grey FA associated proteins are represented, and in different colors the FA proteins. During S-phase of cell cycle, if an ICL is found, the replication fork is stalled. FANCM, together with other FA associated proteins, detects the ICL, activates the DNA damage response mediated by ATR and facilitates the recruitment of the rest of FA core components to the lesion. Two of them, FANCL and FANCT monoubiquitinates the FANCD2-FANCI complex and is recruited to the lesion where recruits FANCQ, FANCP and other proteins to unhook the ICL. Then TLS polymerases repair one of the strands, which serves as DNA template to repair the other strand by HR. Finally when the DNA is repaired USP1 protein deubiquitinates the ID2 complex to finish the process. Modified from Wang and Smogorzewska*[212].

3.2.1. Other important endogenous DNA repair pathways involved in the repair of interstrand-crosslinks

Although FA/BRCA pathway mainly regulates the repair of an ICL in the DNA, other repair pathways are also activated during the ICL repair: nucleotide excision repair (NER) and translesion synthesis (TLS). The activation of these pathways generates a DSB in the DNA that can be finally repaired through HR.

The interaction between the different pathways is highly regulated and is important to maintain the correct homeostasis of the cell.

3.2.1.1. Nucleotide excision repair

Nucleotide excision repair is the preferential via to solve lesions caused by ultraviolet radiation. The common denominator in NER repair is a significant distortion of DNA helix.

The lesion is recognized by the heterodimer formed by XPC and RAD23B; which also acts as a repair-recruitment factor. Then, the transcription factor TFIIH, formed by XPB and XPD, is recruited. These proteins present helicase activity, and therefore mediate strand separation at the side of the lesion. After damage detection, XPA verifies the lesion in an open DNA conformation and RPA protein stabilizes the opened DNA and facilitates the location of the endonucleases XPG and ERCC1-XPF. These endonucleases remove a 27 to 29 base pairs (bp) oligonucleotide containing the lesion and finally the gap is filled by PCNA, δ polymerase and ligases using the complementary strand as a template. These repair is called global genome NER (GG-NER), which is the pathway in charge of repairing the lesion across the entire genome. However, there is another kind of NER repair, called transcription-coupled NER (TC-NER) which repairs transcription-blocking lesions present in transcribed DNA strands and is involved in replication and recombination-independent ICL repair during G0/G1 phases of the cell cycle[196,197]. In TC-NER, DNA damage is detected by elongating RNA polymerase II complex, and the damage is much faster eliminated. CSA and CSB proteins are in this case needed for faster and more efficient repair of transcribed strands[213,214] (Figure 11).

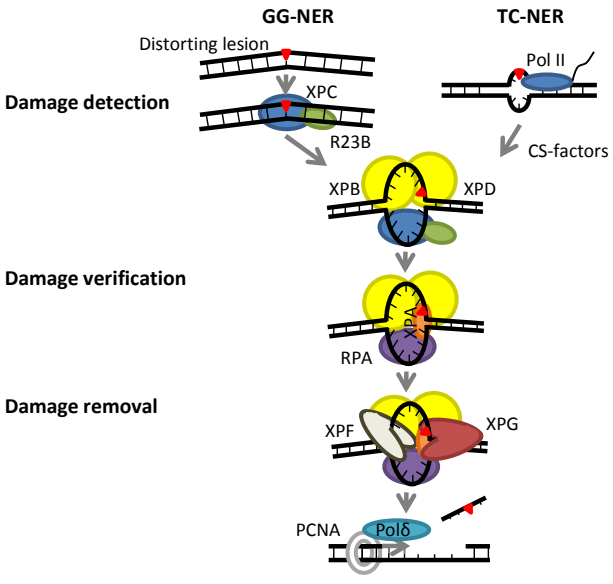


Figure 11: Nucleotide excision repair (NER) pathway *is involved in the removal of damaged bases in the genome. Red triangle represents the distorting base. In GG-NER: global genome NER, damage is*

detected by XPC together with RAD23B. Then XPB and XPD mediate strand separation. The damage is verified by XPA and RPA stabilizes the opened DNA. XPG and XPF endonucleases remove a DNA fragment that contains the lesion and the gap is filled by PCNA, δ polymerase and ligases using the complementary strand as template. In TC-NER: transcription-coupled NER the damage is detected by RNA polymerase II and the CS-factors are needed for faster and repair. Modified from de Laat et al[213] and from Gillet and Schärer[215].

3.2.1.2. Translesion synthesis

Translesion synthesis is a DNA damage tolerance (DDT) pathway that commonly allows replication in the cell across polymerase-blocking lesions. It specifically allows replication machinery to bypass DNA lesions using low-fidelity DNA polymerases. These low-fidelity polymerases are non-processive, lack any proofreading capability and contain larger active sites capable of accommodating distorted bases and base pair mismatches. This is an error-prone pathway, as the lesion bypass occurs at the cost of replication fidelity[216].

Firstly RAD6 (E2 ubiquitin-conjugating enzyme) binds to RAD18 (E3 ubiquitin ligase) in response to DNA damage. These two enzymes induce monoubiquitination of PCNA to facilitate the switch between DNA replication and DDT processes. This monoubiquitination increases PCNA affinity for the low fidelity polymerases at the damaged site. The exchange of replicative DNA polymerase stalled for a TLS polymerase occurs in a multi-step process[217]. Firstly one TLS polymerase, called the insertion polymerase, is recruited by PCNA monoubiquitination. The insertion polymerase incorporates one nucleotide opposite to the damaged site. The insertion polymerase is replaced by an extension TLS polymerase which extends the TLS patch by around 18 nucleotides, avoiding the detection of the lesion by the 3'-5' exonuclease proofreading activity of a replicative DNA polymerase. Finally, there is a switched back to the high fidelity DNA polymerase (Figure 12).

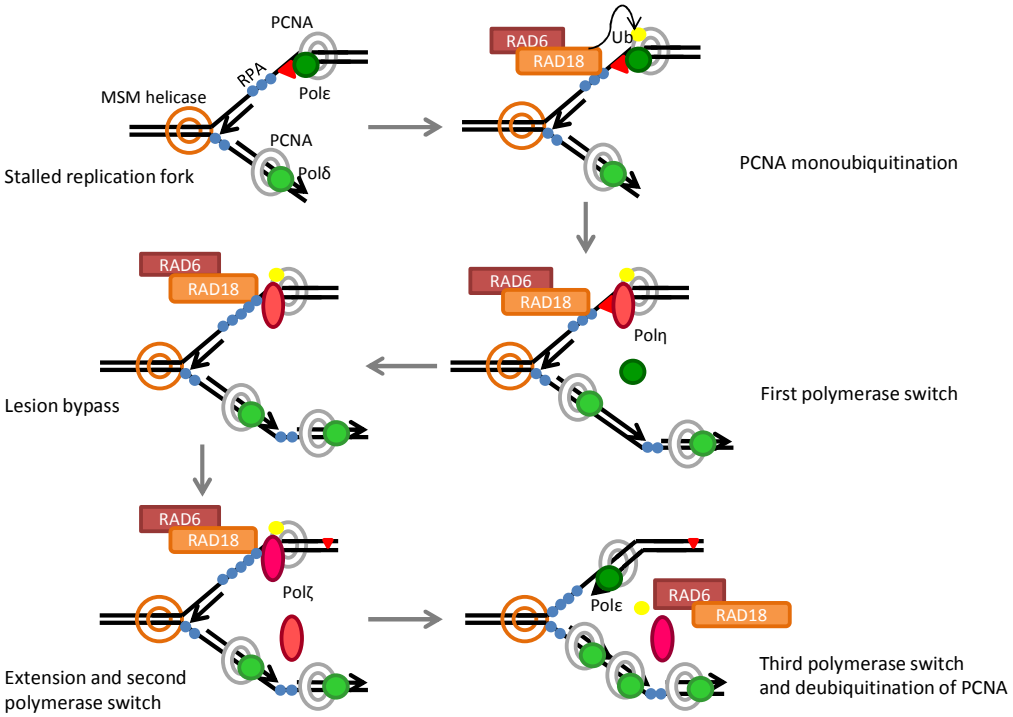


Figure 12: Translesion synthesis (TLS) pathway, *a DNA damage tolerance mechanism involved in replication over modified or damaged DNA bases. Red triangle represents the distorting base. RAD6-*

RAD18 complex monoubiquitinates PCNA, which increments the affinity of this factor to low fidelity polymerases. Then the first polymerase switch occurs, the insertion polymerase is recruited and incorporates one nucleotide opposite to the damage site. This process enables to bypass the lesion. Then the insertion polymerase is replaced by extension polymerase, which extends the patch to avoid the 3'-5' exonuclease proofreading activity of the replicative DNA polymerases. Finally, the third polymerase switch is produced, PCNA is deubiquitinated and the replicative polymerase continues replicating the DNA. Modified from Ghosal and Chen[218].

3.2.2. Interactions between FA/BRCA and other DNA repair pathways

The FA/BRCA pathway is involved in the coordination of different repair mechanisms to solve ICL such as NER, TLS and HR. Moreover, FA/BRCA pathway is also involved in regulation of NHEJ.

A clear example of the close interaction between these pathways is the case of mutations in XPF, a member of NER pathway. Mutations in this gene can provoke three different disorders: xeroderma pigmentosum[219], progeroid syndrome[220] or FA[189]. Moreover, RPA - involved in both HR and FA dependent ICL repair - can be associated to post-incision complexes, preventing further incision events and coupling effectively initiation of NER to completion of DNA repair synthesis[221].

There is also a clear link between TLS and FA/BRCA pathways. RAD6/RAD18 - an important complex in TLS - can directly impact in FA network activation via FANCD2 ubiquitination[222-224]. Moreover, monoubiquitinated PCNA is also able to recruit FANCL to the chromatin[222]. This and other studies suggest that PCNA acts as a scaffold protein for both FANCL and FANCD2 recruitment to the chromatin by promoting FANCD2 ubiquitination. FANCD2 monoubiquitination - with the subsequent activation of downstream proteins - also plays a role in recruiting polymerase η to the sites of damage, potentially facilitating TLS polymerase switching[217]. RAD18 protein is recruited to the ssDNA in a RPA dependent manner process that links TLS and HR[225]. FAAP20 - a FA-associated protein present in the core complex - binds to monoubiquitinated Rev1 (insertion polymerase) and promotes its interaction with the FA core complex[226]. Finally, it is known that monoubiquitinated PCNA is deubiquitinated by USP1 as well as ID2 complex of FA[227], showing that regulation of both pathways should be very similar.

FANCD2 affects all repair mechanisms of DSB when the replication fork is stalled. It is well known that its monoubiquitinated form has a role in DNA repair, promoting HR and single strand annealing (SSA)[228-231]. FA deficient cells present defects in HR when DNA damage is produced through ICLs[232]. On the other hand, monoubiquitinated FANCD2 inhibits non homologous end-joining (NHEJ) dependent on DNA-PKcs[233,234]. Human cell lines deficient in NHEJ proteins are not hypersensitive to ICL agents. However, some studies have shown that NHEJ inhibition in FA cell lines reduces the toxicity of ICL agents. These studies demonstrated that during the ICL repair, FA proteins prevent the inappropriate binding of NHEJ effectors in the DNA ends and suppress the imperfect ligation of DSB induced by ICL between non homologous chromosomes[233,234]. However, the inhibition of NHEJ in FA cells seems to be more deleterious in patient's cells. For example the elimination of 53BP1 or Ku80 in FANCD2 deficient cells increases its genetic instability[235]. FANCD2 has also been involved in the processing of DSBs promoting HR: monoubiquitinated FANCD2 carries CtIP protein to

chromatin to initiate resection of the DSB during the repair process of the ICL. Mutant cells deficient in CtIP domains that bind FANCD2 have increased NHEJ activity, resulting in hypersensitivity to ICL. Therefore, FANCD2 participates in the competition between NHEJ and HR pathways to repair during DSB processing[236,237]. It has been also described a model in which 53BP1 inhibits initiation of DNA end-resection favoring the repair by NHEJ. This inhibition is negatively regulated by FANCS and thus, loss of FANCS results in persistent inhibition of end-resection, which can be abolished by removal of 53BP1[53]. This could at least in part explain the inhibition that FA/BRCA pathway exerts over NHEJ.

FANCM also plays a role in HDR repair process independently of the FA network, as its depletion reduced RPA foci formation at ICL, while depletion of the rest of the FA core does not produce a similar response[238]. It has also been demonstrated that inactivation of FANCG decreases HR[239]. Finally, a link between B-NHEJ and FA has also been described, as depletion of FANCA protein reduces B-NHEJ[240].

3.3. Clinical aspects of the disease

Fanconi anemia is a rare disease with an estimated frequency of 1-3 patients per 500,000 births, and an estimated frequency of heterozygous mutation carriers of 1 per 300. This frequency is increased in highly consanguineous groups such as the Spanish gypsies, in which the frequency of carriers is estimated as 1 per 67 births[241].

Due to the heterogeneity of the disease, its presentation is very diverse. Around 70% of patients show physical abnormalities which can include: skeletal abnormalities of the hips, spine or ribs; radial ray anomalies including hand and arm anomalies, misshapen, missing or extra thumbs or an incompletely developed or missing radius; vertebral scoliosis; skin pigmentation anomalies such as *café-au-lait* spots, hyper- and hypopigmentation; short stature; small head or eyes; kidney and urinary tract problems, including missing or horseshoe kidney; mental retardation or learning disabilities; low birth weight; gastrointestinal difficulties; cardiac abnormalities; small reproductive organs and infertility in males and reduced fertility in females[242].

The most frequent feature of the disease is the bone marrow failure (BMF), which is manifested at a median age of 8 years, being the primary cause of morbidity and mortality in FA patients. Although some years ago the overall survival of these patients was situated around 25 years[151], the more accurate diagnosis, better care and improvement in hematopoietic stem cell transplant (HSCT) imply that the number of FA patients living into their third and fourth decade is increasing[242]. At birth, blood counts are usually normal, but not the number of hematopoietic stem and progenitor cells (HSPCs) which are reduced[243-245]. Macrocytosis is usually the first symptom of the BMF detected in these patients, followed by thrombocytopenia and neutropenia.

In addition to the BMF, the incidence of cancer in FA patients is also increased compared to the normal population. The incidence of acute myeloid leukemia (AML) is 800-fold higher than in the general population, and patients are also prone to develop myelodysplasic syndrome (MDS). Moreover, FA patients have shown a higher risk of developing specific solid tumors,

such as head and neck squamous cell carcinomas (SCCs), esophagus SCCs and vulvar cancer in women (33% of FA patients develop a solid tumor in their fourth decade of life)[246-248].

An important aspect concerning the clinical symptoms of the disease is the observation that some FA patients can undergo somatic mosaicism by means of a spontaneous reversion of the pathogenic mutation, in particular somatic cells. If this reversion takes place in a HSPC, it may confer a proliferation advantage due to its increased genetic stability, leading to the recovery of the peripheral blood (PB) cell counts in patients[5,249-251].

Interestingly, in 2010, a singular phenotype in some FA patients was first described[252]. The blood cells of these patients did not show the FA characteristic G2/M cell cycle arrest after exposure to DNA cross-linkers. Strikingly and in contrast to mosaic FA patients, no reversion in the pathogenic mutations occurred. This phenomenon called attenuation was then associated with almost normal blood cell counts without the reversion of the FA mutation. However, the cost of this attenuated phenotype in patients is a higher risk of malignancy[252].

3.4. Phenotypic characteristics of Fanconi anemia cells

The main characteristics of FA cells are:

a) Hypersensitivity to ICL agents such as MMC or DEB: the treatment with these agents generates a high number of chromosomal aberrations. This characteristic is a hallmark of FA and it is used to diagnose the disease in PB lymphocytes[253].
b) G2/M cell cycle arrest after exposure to ICLs inducing agents: it is a consequence of the accumulation of blocked repair hairpins due to the incapability of these cells to repair the DNA lesions produced by ICL agents[254]. This technique has been very useful in the diagnosis of mosaic patients, since most of them do not present the typical MMC or DEB sensitivity in PB T lymphocytes. The treatment of fibroblasts with MMC or DEB and the G2/M arrest is usually used for the diagnosis[255].
c) Hypersensitivity to oxygen: this is a consequence of the deregulation of reactive oxygen species (ROS) metabolism[256]. This deregulation causes defects in the *ex vivo* growth of cells. Therefore, to facilitate FA cells growth, they must be cultured in hypoxic conditions (oxygen concentration lower than 5%), mimicking the low oxygen concentrations present in most tissues including the bone marrow niche[257].
d) Telomere abnormalities: FA cells present shorter telomeres than healthy related individuals[258], but not as short as cells from dyskeratosis congenita patients (in which mutations in genes involved in maintenance of telomere's structure occur)[259]. Although the telomeres are usually altered in FA patients; the telomerase activity is increased, probably to compensate this alteration[260,261].
e) Cytokine sensitivity: FA cells from patients produce higher levels of pro-apoptotic and pro-inflammatory cytokines such as interferon γ (IFN- γ)[262], tumor necrosis factor α (TNF-α)[263,264] or interleukin 1β (IL-1β)[265]. The overexpression of TNF-α and IFN-β in the bone marrow of FA patients is associated with the hyperactivity of two pathways involved in bone marrow failure and leukemogenesis[266]: p38MAPK[267] and NF-κB[268]. The hematopoietic phenotype of FA patients is at least, in part, attributable to the overproduction of some cytokines to which FA cells are hypersensitive. Some attempts to ameliorate BMF have been developed with the objective of decreasing the levels of

cytokines such as TNF-α[262]. Although the *in vitro* culture of HSPCs from patients is increased in the presence of anti-TNF-α[262,269], a clinical trial conducted in FA patients with an inhibitor of TNF-α did not improve the clinical status of the patients[270].

f) Adhesion and homing defects: they are linked to defects in CDC42 activity[271] - a GTPase that is downregulated by over-expression of TNF-α.

g) Defects in monoubiquitination and mobilization of FANCD2 to DNA repair foci: monoubiquitination of FANCD2 is essential for the recruitment of ICL repair proteins. Cells with mutations in the FA core complex present defects in the monoubiquitination of this protein and this can also be used as a diagnostic marker[251,272]. Monoubiquitination of FANCD2 could be observed in Western blots (WB) by the appearance of two bands. Since monoubiquitination of FANCD2 is essential for FANCD2 foci formation, the analysis of FANCD2 foci can also be an indicator of FANCD2 monoubiquitination[202].

h) Defects in cytokinesis, due to the accumulation of monoubiquitinated ID2 complex during and after mitosis at fragile sites in which DNA replication is slower. Sister chromatids in which these foci are located are connected by ultra-fine bridges. This characteristic increases the number of binucleated or multinucleated cells that arise from cytokinesis failure[273].

i) Elevated levels of p53 protein and its effector p21 in patient cells is characteristic, and this could reflect the unresolved DNA damage of the cells, which might then lead to hematopoietic stem cell (HSC) depletion and hence, progressive BMF[160]. The strong induction of p53 gives rise to a new cell cycle arrest in G0/G1 phases, as a protective mechanism to keep the integrity of the genome, preventing damaged cells to progress through the cell cycle. Some *in vitro* studies have used small molecule inhibitors of p53, resulting in rescue of HSC apoptosis and improving hematopoiesis in FA patient´s cells, but this increases the risk of leukemia[244].

3.5. Diagnosis and identification of complementation groups in Fanconi anemia

The diagnosis of FA patients is not an easy procedure. As it has been previously mentioned, one third of them do not present major malformations, so they are not identified until the first symptoms of the BMF appear[161]. Moreover, there are other diseases that reproduce some of the phenotypes of FA patients such as Nijmegen breakage syndrome (NBS), Bloom syndrome and Seckel syndrome.

The differential test for FA is the chromosomal breakage test in the presence of ICL agents. This was first established by Schuler and colleagues in 1969, who observed that FA patient's cells were positive for MMC chromosomal breakage induction, while healthy donors' cells were negative[274]. The diagnosis of FA can also be done by a G2/M phase cell cycle arrest after treatment with ICL inducing agents[255]; and this can be very useful to confirm cases of mosaicism.

The screening of FA complementation groups is performed to detect which is the mutated gene in a particular patient and assigned him/her to a complementation group. One widely method consists in the transduction of peripheral blood T cells using a battery of vectors with

either γ-retroviral (γ-RVs) or lentiviral vectors (LVs), each one encoding a different FA gene, followed by the analysis of the vector that corrects the hypersensitivity of these cells to interstrand cross-linking agents. The vector that reverts the phenotype determines the FA complementation group of the patient[275].

If the patient does not have mutations in the most common FA genes, WB of FANCD2 allows defining if the FA affected protein is located upstream or downstream of the ID-complex in the FA pathway, depending on the ubiquitinated status of the protein.

Nowadays new approaches based on exome sequencing through the analysis of the 19 FA genes described so far, or the analysis of the entire genome, can allow the identification of new complementation groups, as in the studies that facilitate the of identification of FANCQ and FANCR as a FA genes[189,190]. In the near future this kind of analysis will substitute the subtyping carried out with γ-RVs and LVs. However, subtyping will continue to be essential to verify whether the insertion of the suspected FA gene restores the phenotype of the cell, confirming the pathogenicity of the mutations found by sequencing.

3.6. Current treatments in Fanconi anemia

Bone marrow failure (BMF) is the main cause of mortality in FA patients. Many of the treatments to avoid the BMF are palliative, and directed to maintain acceptable numbers of cells in PB. The standard therapies are based on the use of androgens, hematopoietic growth factors[276], and if possible HSCT, that is the only curative treatment. Androgens stimulates erythropoiesis and are recommended when residual endogenous hematopoiesis remains[277], however this treatment presents side effects such as masculinization[278] or higher risk of hepatic tumors[279]. Recent results using danazol have shown that this androgen can be an effective and well-tolerated treatment option to delay the progressive marrow failure in FA patients and with less side effects than the previous androgens used[280]. Some hematopoietic growth factors that have been also used in FA treatment are granulocyte-colony stimulating factor (G-CSF), erythropoietin (EPO) or interleukin 3 (IL-3)[281,282]. However, the response to these factors was transient and they are no longer used.

The only curative treatment capable of restoring the hematopoiesis of FA patients in the long-term is allogenic HSCT. In 1984, Gluckman and coworkers[283] developed the first successful preparative regimen for FA patients. However, there are many obstacles for HSCT for example the hypersensitivity of FA patients to conditioning regimens or the finding of a compatible donor. The ideal donor for a FA patient is an HLA-identical sibling donor, to limit the risks of graft rejection and graft versus host disease (GVHD). Nevertheless, recent advances including the depletion of alloreactive T cells, the use of reduced intensity conditioning and the addition of fludarabine (potent inmunosupressor which interferes with DNA synthesis) have also shown to improve the outcome of alternative (HLA-matched related or unrelated) donor transplants in FA patients[284,285]. However and although HSCT is the primary treatment of these patients, it does not solve the non-hematological complications of FA, particularly the increased incidence of solid tumors.

The management of cancer in these patients is different than in other patients due to the toxicity of chemo- and radiotherapeutic regimens. Prevention is the critical element, so FA

patients should have annual analysis of bone marrow aspirates to rule out premalignant clonal expansion. Also frequent head and neck evaluations, and gynecologic exams in the case of women, are strongly recommended to prevent SCC. It is also recommended that all FA patients should receive human papillomavirus (HPV) vaccinations to prevent SCCs[286].

3.6.1. Gene therapy in Fanconi anemia

Fanconi anemia is a monogenic disease and a good candidate for GT due to the proliferative advantage of corrected cells over non-corrected ones[5]. This phenomenon is naturally produced in mosaic patients, where the reversion of the pathogenic mutation in one cell leads to the correction of the hematological status of these patients. However, FA also presents some disadvantages in the field of GT; the main one is the difficulty to collect high numbers of HSCs from patients, as the pool of available HSPCs is already compromised at birth[244].

Despite all of this, three different gene therapy trials have been carried out using γ-RVs in patients with FA. The first one was carried out by Dr. Walsh in FA-A patients, but was not successful[287]. The second one was conducted by Liu and colleagues[288] in FA-C patients. In this study the authors observed an increased number of BM cells in treated patients, but no clinical improvement was observed. The third study was developed by Kelly et al. to treat patients from FA-A complementation group. Neither, this study generated clinical improvements in the patients[243].

The lack of success observed in previous FA GT trials gather in 2011 worldwide experts in the field to create an International FA GT working group, in order to set up the conditions for future FA GT trials[289]. In 2012 a phase I clinical trial in patients of the complementation group A was opened at the Fred Hutchinson Center in Seattle, USA (NCT01331018). The first patient was recruited in 2014 and results have not been published yet. Our laboratory coordinates FA GT trial in Europe. This trial uses a SIN LV vector[290] which was designated as an orphan drug in 2010 by the European Commission (EMA/COMP/662962/2010) and was approved for phase I/II GT trial for FA-A patients (FANCOLEN: Eudra number CT: 2011-006100-12), the most frequent FA complementation group around world (>60%) and also in Spain (>80%)[275].

3.6.1.1. Gene targeting in Fanconi anemia

Gene targeting will represent the safest gene therapy alternative to treat monogenic disorders, as extensively mentioned in section 2 of the introduction. Although nowadays there is no clinical trial in this respect to treat FA, some experimental studies have already addressed this topic in FA cells. The first work was reported by our laboratory. In this study, fibroblasts from FA-A patients were edited with AdVs which contained ZFNs that targeted the *AAVS1* locus and a donor IDLV which carried that *FANCA* gene. Corrected cells were then reprogramed to generate induced pluripotent stem cells (iPSCs) and finally differentiated to the hematopoietic lineage[75]. In the next study, fibroblasts from FA-A patients were directly reprogramed to iPSCs and iPSCs were then gene edited with helper-dependent AdVs in a knock-in strategy, to finally differentiate to the hematopoietic lineage[291]. The use of helper-dependent AdVs without using nucleases markedly decreased the efficiency of gene targeting, compared to the previous work. The main limitation of using reprograming approaches is that the number of hematopoietic stem cells obtained after differentiation from iPSCs is very limited. The last work was based on the correction of FA-C fibroblasts using CRISPR/Cas9

system[292]. In this work CRISPR/Cas9 nucleases and nickases were used with the donor as plasmid DNA in a knock-in strategy. The main conclusion of this work was that although the nickase reduced of NHEJ events to repair the DSBs, it also decreases the efficiency of HDR with respect to the nuclease.

So far, no gene editing studies have been performed in hematopoietic cells from FA patients, which constitute the main target in FA gene therapy. For this reason in the present work we have addressed this topic and we have investigated different options to target the hematopoietic cells, and thus to correct the hematopoietic phenotype of FA-A patients.

OBJECTIVES

The aim of this study was to investigate the feasibility of developing a targeted gene therapy approach in hematopoietic stem and progenitor cells from Fanconi anemia patients, subtype A (FA-A HSCs). In particular, our approach was focused on the specific insertion of the therapeutic *FANCA* gene into the *AAVS1* safe harbor locus of FA-A HSCs.

To achieve this objective, the following specific aims have been pursued:

To construct a self-inactivating integrase-defective lentiviral vector harboring the *FANCA* therapeutic gene, as donor construct to be inserted into the *AAVS1* locus, targeted with a specific engineered zinc-finger nuclease (ZFN).

To investigate the feasibility, efficacy and specificity of a gene targeting approach aiming at the insertion of reporter and FA therapeutic genes into the *AAVS1* safe harbor locus of lymphoblastic cell lines and hematopoietic stem cells from healthy donors.

To improve the efficacy of gene targeting by increasing the transduction efficacy of integrase-defective lentiviral vectors and enhancing the homologous recombination frequency in FA-A cells.

To evaluate the efficacy of the proposed gene targeting approach to correct the phenotype of lymphoblastic cell lines and hematopoietic stem and progenitor cells from FA-A patients.

MATERIALS AND METHODS

1. Cell lines and primary cells

1.1. 293T cell line

293T cells (ATCC: CRL-11268) are used to produce non replicative LVs. It is a human embryonic kidney epithelial cell line that was modified by the insertion of a temperature sensitive gene codified by the simian SV40 T antigen that is constitutively expressed to achieve higher transfection efficiency. The inserted sequence allows the episomal replication of plasmids that contain the replication origin in the early promoter region of the SV40.

1.2. Lymphoblastic cell lines

Healthy donor or FA patients' B cells from peripheral blood were transformed with the Epstein-Barr virus (EBV) that activates the proliferation of these cells. These resulting lymphoblastic cells (LCLs) present a normal karyotype, are immortalized and are not tumorigenic[293]. Four different cell lines from healthy donors were used in this study CP1, CP2, C3 and C4, and four cell lines from Fanconi anemia A complementation group (FA-A) patients (FA-55, FA-56, FA-122, FA-378).

1.3. Human fibroblasts

Fibroblasts obtained from healthy donors and also from FA-A complementation group patients.

1.4. $CD34^+$ cells derived from umbilical cord blood

Umbilical cord blood (UCB) samples from healthy donors were obtained from the Centro de Transfusiones de la Comunidad de Madrid. Mononuclear cells were purified by Ficoll-Paque PLUS (GE Healthcare, Fairfield, USA) density by gradient centrifugation. Then, $CD34^+$ cells were selected using CD34 MicroBead Kit. Magnetic-labelled cells were selected with a LS column in QuadroMACS™ separator (MACS, Miltenyi Biotec, Bergisch Gladbach, Germany) following manufacturer's instructions. Purified $CD34^+$ were then evaluated for their purity by flow cytometry. Purities from 90-99% were routinely obtained.

1.5. $CD34^+$ from FA patients

A very small proportion of mobilized peripheral blood (mPB) $CD34^+$ cells that remained in cell collection bags and tubes from the CliniMACS® System (Miltenyi Biotec), was used. This system is used to purify $CD34^+$ cells from FA-A patients treated with plerixafor and G-CSF included in the FANCOSTEM mobilization trial. This procedure was performed after the informed consent of patients or their parents and with the approval of the ethic committee of the hospitals.

1.6. Culture conditions

The culture conditions of the different cell types used in this study are listed in Table 3:

293T	**Culture medium**	Iscove's Modified Dulbecco's Medium (IMDM, Gibco/Life Technologies/Thermo Fisher Scientific, Waltham, USA)) + 10% Hyclone (GE Healthcare) + 1% Penicillin/Streptomycin (P/S, Gibco)
	Growth	Adherence
	Culture conditions	Normoxia: 37°C, 21% O_2, 5% CO_2 and 95% relative humidity (RH)
LCLs	**Culture medium**	Roswell Park Memorial Institute (RPMI, Invitrogen/Life Technologies/Thermo Fisher Scientific, Waltham, USA) + 20% Hyclone + 1%P/S + 0.005 mM β-Mercaptoethanol (Gibco) + 1mM sodium pyruvate (Sigma, Saint Louis, MO, USA) + non-essential aminoacids (NEEA, Lonza, Basel, Switzerland)
	Growth	Suspension
	Culture conditions	Normoxia: 37°C, 21% O_2, 5% CO_2 and 95% RH
Human fibroblasts from healthy donors	**Culture medium**	*Dulbecco's Modified Eagle's medium* (DMEM, Invitrogen) + 20% Hyclone + 1% P/S
	Growth	Adherence
	Culture conditions	Normoxia: 37°C, 21% O_2, 5% CO_2 and 95% RH or Hypoxia: 37°C, 5% O_2, 5% CO_2 and 95% RH when these cells are used in experiments together with fibroblasts from FA-A patients
Human fibroblasts from FA-A patients	**Culture medium**	DMEM + 20% Hyclone + 1% P/S
	Growth	Adherence
	Culture conditions	Hypoxia: 37°C, 5% O_2, 5% CO_2 and 95% RH
UCB $CD34^+$	**Culture medium**	StemSpam (StemCell Technologies, Vancouver, Canada) + 1% P/S + 1% Glutamax (Gibco) + 100 ng/mL stem cell factor (SCF, EuroBiosciences, Friesoythe, Germany) + 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3, EuroBiosciences) + 20 ng/mL thrombopoietin (TPO, EuroBiosciences) + 20 ng/mL interleukin 6 (IL6, EuroBiosciences)
	Growth	Suspension
	Culture conditions	Normoxia: 37°C, 21% O_2, 5% CO_2 and 95% RH
$CD34^+$ from FA patients	**Culture medium**	StemSpam + 1% P/S + 1% Glutamax + 100 ng/mL SCF + 100 ng/mL Flt3 + 20 ng/mL TPO + 20 ng/mL IL6 + 10 µg/mL anti-TNFα (Enbrel-Etanercept, Pfizer, New York, USA) + 1 mM N-acetylcysteine (Pharmazam, Spain)
	Growth	Suspension
	Culture conditions	Hypoxia: 37°C, 5% O_2, 5% CO_2 and 95% RH

Table 3: Culture conditions of the different cell types used in this study.

2. Lentiviral vectors and plasmids

Bacterial plasmids containing the different constructs used for LV production were generated using typical procedures of molecular biology. All enzymes used during the cloning such as restriction enzymes, ligases, kinases and phosphatases were obtained from New England Biolabs (Ipswich, USA). After each ligation, the resulting products were transformed in TOP10 bacteria (Invitrogen). Plasmid DNA was purified from selected colonies using NucleoBond® Xtra Midi EF kit (Macherey-Nagel, Duren, Germany) and analyzed by restriction enzymes.

2.1. Zinc-finger plasmids

Specific ZFNs to target the *AAVS1* locus are included in the pVAX plasmid (pVAX ZFN) or pUC vector (TEV ZFN). ZFNs were designed by Sangamo Bioscience (Richmond, California, USA) and kindly provided by Pr. Luigi Naldini (Hospital San Raffaelle, Milano, Italy). Both vectors contain the same sequence: the homology arm (HA) for intron 1 of *PPP1R12C* gene and a mutated version of *Fok*I nuclease that increases the affinity by forced heterodimerization[101,102], for this reason two different plasmids are required. One member of the pair presents the right homology arm (HA R) for the *AAVS1* locus and *Fok*I with EL mutation (Q486E and I499L) and the other member presents the left homology arm (HA L) for the *AAVS1* locus and *Fok*I with KK mutation (E490K and I538K). The difference between pVAX and pTEV is that TEV ZFNs pair presents regulatory sequences that enhances the stability of the *in vitro* produced mRNA whereas the pVAX ZFNs pair does not.

These plasmids have been used in all the experiments performed with ZFNs as mRNA and this mRNA was synthesized *in vitro* (see section 3. Materials and methods).

2.1.1. pVAX zinc-finger nucleases

We used two different plasmids, one for the right ZFN and other for the left ZFN. In both plasmids there is a T7 promoter that drives the transcription of the ZFN from them followed by a nuclear localization signal (NLS), the HA right or left and finally, the mutated version of the nuclease *Fok*I. The obligate heterodimer *Fok*I domains generate an active nuclease only by heterodimerization and when incorporated into ZFNs induce DSBs with higher specificity. After the nuclease sequence a unique restriction enzyme site is included to allow linearization of the plasmid before the *in vitro* transcription.

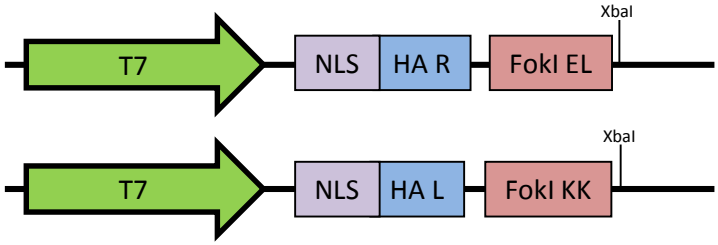


Figure 13: Schematic representation of pVAX ZFNs. *T7 promoter: from which the mRNA is* in vitro *synthetized. NLS: nuclear localization signal. HA: homology arms (R: right or L: left) of PPP1R12C gene located in the AAVS1 locus. Obligated heterodimer FokI domain (FokI EL or FokI KK). XbaI indicates the unique restriction site present that allows linearization of the vector to improve the efficiency of the* in vitro *synthesis.*

2.1.2. TEV zinc-finger nucleases

There are two different plasmids, one that encodes the right ZFN and the other that encodes the left one. The transcription of these two plasmids is also under the control of the T7 promoter. The plasmid also carries a NLS, the homology arm for the *PPP1R12C* gene and the mutated version of the *Fok*I endonuclease. In this case and in contrast with the pVAX vector, it also has two different regulatory elements that presumably confer more stability to the mRNA[294]. The 5' untranslated region (UTR) comes from the tobacco etch virus (TEV) and the 3' UTR comes from the human β-globin gene.

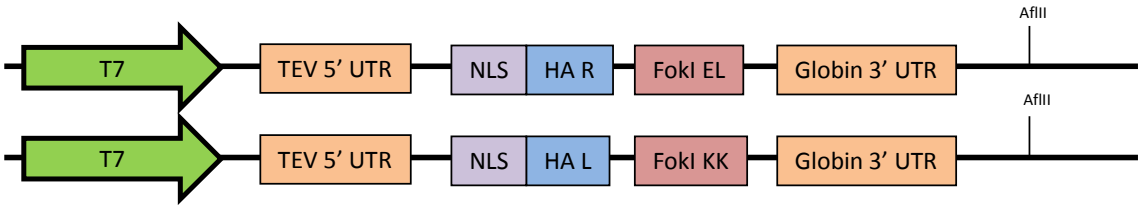


Figure 14: Schematic representation of TEV ZFNs. *T7 promoter: from which the mRNA is in vitro synthetized. NLS: nuclear localization signal. HA: homology arms (R: right or L: left) of PPP1R12C gene in the AAVS1 locus. Regulatory sequences: TEV (tobacco etch virus) in 5' UTR (untranslated region) and β-Globin in 3' UTR. Obligated heterodimer FokI domain (FokI EL or FokI KK). AflII indicates the unique restriction site present that allows linearization of the vector and consequently its in vitro synthesis.*

2.2. Donor constructs

All the donor constructs used in this study are self-inactivating lentiviral vectors (SIN-LV) and were produced as integrase-defective lentiviral vectors (IDLV) to prevent their integration in the genome[295]. All of them present the same homology arms for the *AAVS1* locus described previously[65], but the recombination cassette between the homology arms differs. All the constructs were used as donor template as IDLV or plasmids for the homologous recombination during gene targeting experiments.

2.2.1. GFP donor

This plasmid has been kindly provided by Dr A. Lombardo[65] (Hospital San Raffaelle) and includes the physiological human promoter: phosphoglycerate kinase (PGK) that drives the expression of a reporter gene: green fluorescent protein (*GFP*) flanked by the homology arms (HAs) for the *AAVS1* locus. This construct was also produced as integrase-competent lentiviral vector (ICLV) in the experiments of improvement of gene targeting using gold nanoparticles (AuNPs).

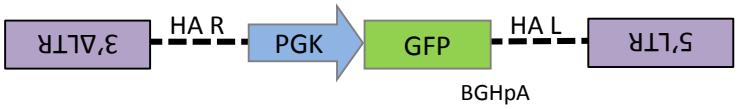


Figure 15: Schematic representation of GFP donor. *The HDR cassette is included between both HAs. All the construct is flanked by the LTRs.*

2.2.2. GFP-FANCA donor

This donor plasmid was previously developed in our laboratory[75]. The construct contains HA for *AAVS1* locus and a promoterless reporter gen (*GFP*) whose expression is driven by the endogenous promoter of the *PPP1R12C* gene through the self-cleaving 2A peptide sequence[296] introduced before the reporter gene. It also contains PGK promoter that controls the expression of the human *FANCA* gene.

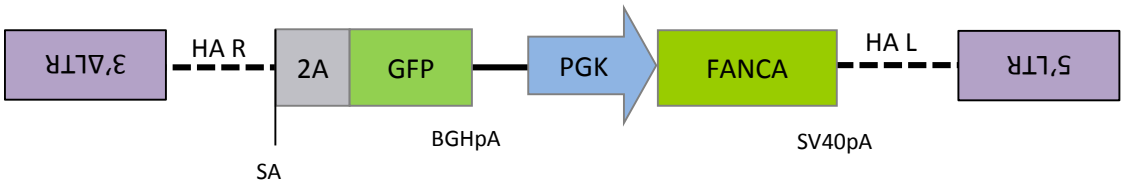


Figure 16: Schematic representation of the GFP-FANCA donor. *Between the 3' and 5' LTRs this plasmid includes two HA for the AAVS1 locus, a splicing acceptor site (SA), the reporter gene GFP, the human promoter PGK and the FANCA gene.*

2.2.3. FANCA-Puro donor

We have designed this plasmid as an alternative donor for gene targeting experiments. It carries the HA for the *PPP1R12C* gene in the *AAVS1* locus, the physiological promoter PGK driving the expression of *FANCA* gene, and the puromycin resistance gene, as a selectable marker. It also contains a 3xFlag signaling peptide before the *FANCA* gene.

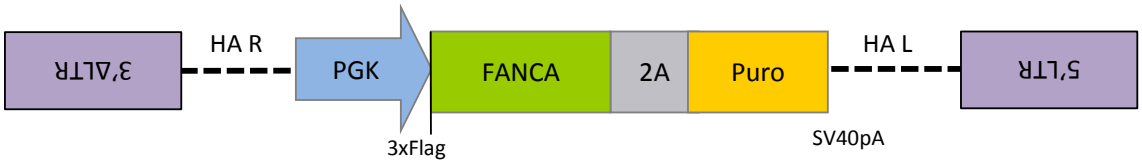


Figure 17: Schematic representation of the FANCA-Puro donor. *Between the 5' and 3' LTRs this construct includes the two HA for the AAVS1 locus, the PGK promoter that drives the expression of the FANCA gene and the Puromycin resistance gene as a selectable marker. Finally, this plasmid also contains the 3X Flag signaling peptide at 5' position of the FANCA gene.*

2.2.3.1. Cloning strategy

The 3xFlag-FANCA-2A-Puro structure was chemically synthetized by GenScript (Piscataway Township, USA) in a pUC57 backbone. 3xFlag signaling peptide was added to the N-termini of the protein as it has been described that in this position it does not produce any alteration in the protein function or localization[297]. The cloning is summarized in the supplementary figure 1 (appendix I). Firstly, a multicloning site was introduced in the plasmid: *Xba*I-*Nhe*I-*Age*I-*Hpa*I.

XbaI NheI AgeI HpaI
5' CTAGAAGCGCTAGCGCCACCGGTCGCAGTT 3'
3' TTCGCGATCGCGGTGGCCAGCGTCAA 5'

Figure 18: Primer annealing of multicloning site *designed to introduce the restriction enzyme sites for NheI and AgeI. In different colors are highlighted the different restriction sites for the different enzymes.*

Two different primers were designed (Figure 18); they were combined at equal concentrations and annealed by incubation 4 minutes at 95°C followed by 10 minutes at 70°C in the annealing buffer (100 mM potassium acetate, 30 mM HEPES pH 7.4 and 2 mM magnesium acetate). They were cooled down at room temperature (RT) and phosphorylated using a PNK kinase and T4 ligase buffer 30 minutes at 37°C followed by 10 minutes at 70°C. At the same time the pUC57 3xFlag-FANCA-2A-Puro plasmid (8,032 bp) was digested with *Xba*I and *Hpa*I restriction enzymes and dephosphorylated with CIP enzyme during 1 hour at 37°C followed by 10 minutes

at 70°C, in order to inactivate it. Annealed primers and the resulting fragment from pUC57 3xFlag-FANCA-2A-Puro digestion were ligated with T4 DNA ligase and transformed in TOP10 bacteria to obtain the first intermediate of the cloning (8,017 bp).

This first intermediate plasmid (8,017 bp) was digested with *Nhe*I and *Age*I restriction enzymes generating a fragment of 8,008 bp. On the other hand, the GFP donor plasmid (9,739 bp) was also digested with the same pair of restriction enzymes generating a fragment of 741 bp. These two resulting fragments were ligated with T4 DNA ligase and transformed in TOP10 bacteria to obtain the second intermediate of the cloning (8,749 bp).

Finally, this last intermediate was digested with *Nhe*I and *Bsa*BI restriction enzymes and a fragment of 6,019 bp was purified. At the same time the GFP-FANCA donor plasmid (14,642 bp) was digested with the same restriction enzymes leading to a fragment of 8,081 bp. These two resulting fragments were ligated with T4 DNA ligase and transformed in TOP10 bacteria to finally obtain the FANCA-Puro donor plasmid.

3. mRNA synthesis

ZFN plasmids were digested with the mentioned restriction enzymes in order to linearize the plasmids. This digestion was purified using phenol-chloroform and the amount of DNA was measured using a Nanodrop®. The *in vitro* synthesis of the mRNA was performed with the 5X MegaScript Kit T7 Kit (Ambion/Life Technologies/Thermo Fisher Scientific, Waltham, USA) and 3´-0-Me-m7G(5')ppp(5')G RNA Cap Structure Analog (ARCA) (New England Biolabs). Firstly 6 µg of the digested DNA was mixed with 32 µL of dNTPs, 12 µL of ARCA, 8 µL of T7 Enzyme, 8 µL of T7 Buffer 10X and up to 80 µL with nuclease free water and incubated two hours at 37°C. Then 4 µL of Turbo DNase were added to the reaction and incubated again for 30 minutes at 37°C.

Then the polyadenylation reaction was performed in the synthetized RNA using the Poly(A) Tailing Kit (Ambion). 144 µL of nuclease free water, 80 µL EPAP buffer 5X, 40 µL $MnCl_2$ 25 mM, 40 µL ATP 10 mM and 16 µL EPAP enzyme was added to the RNA reaction synthesis, mixed and incubated during one hour at 37°C.

The resulting mRNA was purified using RNeasy Plus Mini (Qiagen, Venlo, Limburg, Netherlands) following the manufacturer's protocol.

Finally, the amount of mRNA was measured with Nanodrop®, aliquoted and stored at -80°C. An aliquot of the synthetized mRNA was run in an agarose gel in denaturing conditions (with formaldehyde) in order to confirm the integrity of the mRNA before and after the polyadenylation reaction (Figure 19).

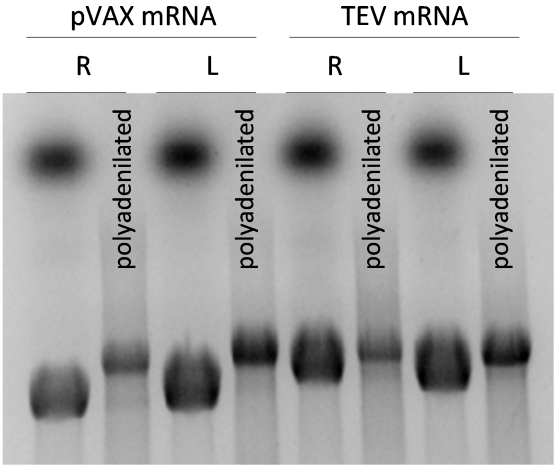


Figure 19: A representative gel of* in vitro *synthetized mRNA from ZFN plasmids: pVAX or TEV. *From each right (R) or left (L) ZFN the non-polyadenilated or polyadenilated mRNA from the L and R ZFN of both constructs is shown.*

Each *in vitro* ZFN mRNA production was tested using a control LCL (C4) in which $2x10^5$ cells were nucleofected with serial dilutions of each ZFN pair using the SF Cell line 4D-Nucleofector® X Kit (Lonza). 24 hours after nucleofection, survival was tested in these cells and five days after nucleofection cells were harvested and the activity of the ZFN pair production was evaluated by Surveyor assay (see section 6. Materials and methods) to determine the optimal concentration of ZFNs to be used in the different productions.

4. Production of integrase-defective lentiviral vectors and integration-competent lentiviral vectors

All self-inactivating (SIN) integrase-defective (IDLV) or integrase-competent (ICLV) vectors were generated by a third generation packaging system in which 293T cells were transiently transfected with the transfer, helpers (pRSV.REV and pMD.Lg/pRRE.D64Vint for IDLV or pMD.Lg/pRRE for ICLV) and envelope (pMD2.VGVg) plasmids (Table 4), obtaining VSV-G-pseudotyped lentiviral particles. The pMD2.VSVg and the helper pRSV.REV plasmids used were obtained from PlasmidFactory (Bielefeld, Germany).

Plasmid type	Name	Description
Packaging plasmids	pMD.Lg/pRRE.D64Vint	Carries *gag* and *pol* genes. The integrase presents a deletion that impairs its expression [295]
	pMD.Lg/pRRE	Comprises *gag* and *pol* genes[298]
	pRSV.REV	Contains *rev* gene[298]
Envelope plasmid	pMD2.VGVg	Presents the glycoprotein G of Vesicular stomatitis virus (*VSV*)[299,300]

Table 4: List of plasmids used in the third generation of IDLV and ICLV production.

Transfections were conducted in 293T cells at 50-70% confluence in 150 mm diameter plates following the $CaCl_2$ DNA precipitation method. Culture medium was replaced with fresh media two hours before transfection. The amounts of plasmids used for a 150 mm plate of 293T cells were: 36 µg of the corresponding transfer plasmid, 9 µg of the pMD2.VSV.G envelope plasmid, 12.5 µg of the pMD.Lg/pRRE.D64Vint or pMD.Lg/pRRE helper, 6.25 µg of the pRSV.REV plasmid and 15 µg of pAdVantage plasmid (Promega, Fitchburg, Wisconsin, United States). The pAdVantage plasmid is described that enhances transient protein expression by increasing translation initiation[301]. This mixture was prepared in a final volume of 1,100 µL of 0.1x Tris-EDTA buffer/dH_2O (2:1) per plate and then 150 µL of 2.5 M $CaCl_2$ was added. After 15 minutes of incubation at room temperature (RT) in agitation in order to allow the correct homogenization of the mixture; 1,250 µL of 2X HBS buffer (100 mM HEPES, 281 mM NaCl, 1.5 mM Na_2HPO_4, pH 7.15) were added dropwise while vortexing at full speed, allowing the formation of Ca^{2+}/DNA^- precipitates. Immediately, the total volume was added to 293T cells, which will subsequently phagocyte the precipitates. After 13 hours, culture medium was replaced by fresh medium. Lentiviral supernatants were collected 36 hours post-transfection, filtered through 0.2 µm pore-size filters (Milipore/Merck KGaA, Darmstadt, Germany) and concentrated by ultracentrifugation. Viral pellets were then resuspended in StemSpam medium to concentrate them 500 times, aliquoted and stored at -80°C.

4.1. Titration of integrase-defective lentiviral vectors

Viral titers were determined by transduction of 293T cells with serial dilutions of the supernatants. $7.5x10^4$ cells/well were seeded in 6-well tissue culture plates the day before. The same day of the titration, the number of cells per well was determined. Serial dilutions of the IDLV supernatants were prepared in IMDM-based complete medium starting from 10^{-3} to 10^{-7} and then used to transduce 293T cells. After 3 days, cells were collected.

The titration was performed by quantitative PCR (Q-PCR). DNA was obtained from the transduced 293T cells using NucleoSpin® Tissue (Macherey-Nagel) and then vector copy number (VCN) was analyzed. Q-PCR amplification was performed using specific primers (Table 5) in a 7500 fast real-time PCR system (Applied Biosystems/Life Technologies/Thermo Fisher Scientific, Waltham, USA) using SYBRGreen FastStart Universal SYBR Green Master PCR master mix (Roche, Basel, Switzerland). After 95°C 20 seconds incubation, amplification was performed for 40 consecutive cycles at 95°C 3 seconds and 60°C 30 seconds. Primers were acquired from Grupo Taper (Alcobendas, Spain).

Primer	Sequence (5' to 3')	Tm (°C)
hAlb Fw	GCTGTCATCTCTTGTGGGCTG	64
hAlb Rv	ACTCATGGGAGCTGCTGGTTC	63
ΔU3 Fw	TCACTCCCAACGAAGACAAGATC	63
PBS Rv	GAGTCCTGCGTCGAGAGAG	62

Table 5: List of primers used for the Q-PCR for titration of IDLVs.

The VCN is the parameter that indicates the number of infective proviral copies per transduced cell. Proviral genome and host cell genome were amplified and quantified using fast SYBR green based multiplex Q-PCR. In this reaction primers that bind the retrotrascribed DNA of the virus were used to quantify viral genome load and human albumin (hAlb) primers were used to

quantify human genome (Table 5). One of the primers that amplified the viral genome binds to ΔU3 and the other binds to the primer binding site (PBS) region. These regions are separated in the RNA viral genome however, once retrotranscription occurs in the cytoplasm of the transduced cell, and the LTRs are formed these two regions become together and the amplification can take place (Figure 20).

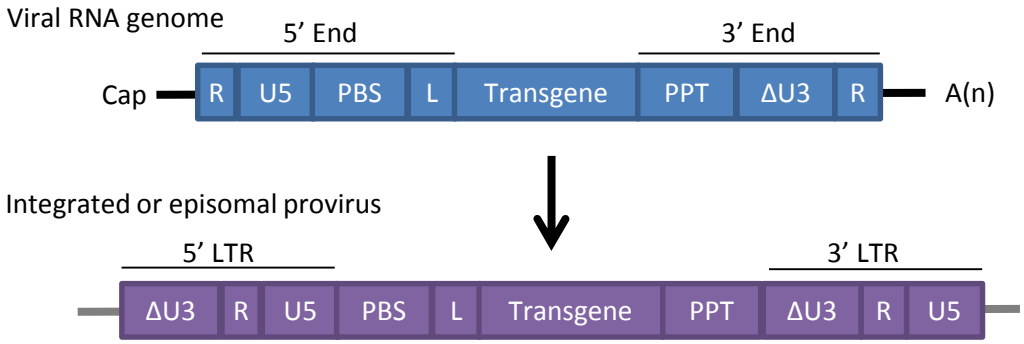


Figure 20: Scheme of the RNA genome of a lentiviral vector and the retrotranscribed form of the vector. *Only when the vector is retrotranscribed the ΔU3 and the PBS regions are close enough to allow the amplification by Q-PCR.*

Quantification of viral and cell genomes was performed by interpolating the cycle threshold values (Ct) from the gDNA samples to the Ct of a standard curve with five-fold dilutions of a known concentration of gDNA from a cell line that presents only one integration of a LV per cell which express GFP and Albumin (kindly provided by Dr. Anne Galy (Genethon, Evry, France))[302]. Cell number corresponding to the calculated number of hAlb copies was determined assuming that each cell contains two copies of hAlb. VCN was finally calculated as:

$$VCN\ (proviral\ copies/cell) = \frac{number\ of\ proviral\ copies}{\left(\frac{number\ of\ Alb\ copies}{2}\right)}$$

The viral titer was calculated for each dilution and then an average of them was used as the final titer expressed in viral particles (vp) per mL. The formula used for titer calculation is the following:

$$Viral\ titer\ \left({}^{vp}/_{mL}\right) = Cell\ number\ at\ the\ day\ of\ transduction \times VCN \times Dilution\ Factor$$

4.2. Titration of integration-competent lentiviral vectors

Viral titers were determined by transduction of 293T cells with serial dilutions of the supernatants. $7.5x10^4$ cells/well were seeded in 6-well tissue culture plates the day before. The same day of the titration, cell number in each well was determined. Serial dilutions of the IDLV supernatants were prepared in IMDM-based complete medium starting from 10^{-3} to 10^{-7} and then used to transduce 293T cells. After 10-15 days, cells were collected.

The titration was performed as described with the IDLV vectors in the section 4.1. Materials and methods.

5. Gene editing protocol

5.1. Gene editing in lymphoblastic cell lines

5.1.1. Using donor as plasmid

$2x10^5$ LCLs cells from healthy donors or FA patients were nucleofected with 2 µg of each donor plasmid and half of the conditions also with 1.5 µg of total *in vitro* synthetized mRNA (0.75 µg of each ZFN pair) using the SF Cell Line 4D-Nucleofector® X Kit (Lonza). Ten days post nucleofection cells were harvested to conduct: flow cytometry, to analyze the percentage of GFP positive cells; integration analysis in the bulk population by PCR (see section 7.1. Materials and methods) and also the nuclease functional assay (see section 6. Materials and methods).

5.1.2. Using donor as integrase-defective lentiviral vector

Gene editing protocol for LCLs is shown in Figure 21. Briefly, $2x10^5$ LCLs cells from healthy donors or FA patients were transduced with the corresponding IDLV at a MOI of 50 in a 96-well culture plate in a final volume of 100 µL. One day after transduction, cells were collected, washed in order to eliminate the viral particles and nucleofected with 1.5-6 µg of total *in vitro* synthetized mRNA (0.75-3 µg of each ZFN pair) depending on the ZFN production, using the SF Cell Line 4D-Nucleofector® X Kit (Lonza). Ten days post nucleofection (day 12 of the protocol) cells were harvested to conduct: flow cytometry to analyze the percentage of GFP positive cells; integration analysis in the bulk population by PCR (see section 7.1. Materials and methods) and also the nuclease functional assay (see section 6. Materials and methods).

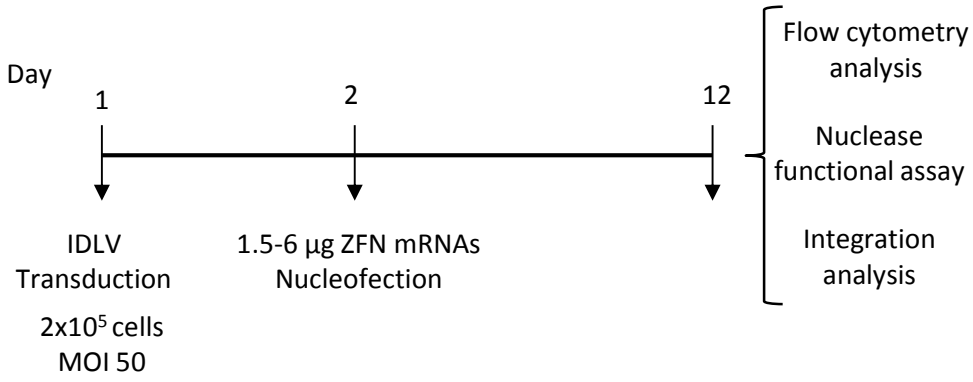


Figure 21: Schematic representation of the gene editing protocol in LCL cells.

5.2. Gene editing in $CD34^+$ cells from healthy donors and Fanconi anemia patients

Gene editing protocol used in UCB $CD34^+$ is shown in Figure 22. Briefly, purified $CD34^+$ cells from healthy donor UCB or FA patients were prestimulated during 48 or 24 hours respectively, with corresponding hematopoietic cytokines at a density of 10^6 cells/mL. Then, 10^5 cells were transduced with the corresponding IDLV at a MOI of 100 in a 96-well culture plate in a final volume of 100 µL. One day after transduction cells were collected and washed in order to eliminate the viral particles and they were nucleofected with 6 µg of total *in vitro* synthetized mRNA (3 µg of each ZFN pair) using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza). 24 hours

after nucleofection 900 cells from healthy donors or 10,000 cells from FA-A patients per M35 plate were seeded in methylcellulose (StemMACS™ HSC-CFU complete with Epo, Miltenyi biotech) to perform the clonogenic assays (see section 5.2.2. Materials and methods). Three days post nucleofection (at day 7 of the protocol) the first flow cytometry analysis was performed as described in the section 5.2.1. Materials and methods. At day 11, 10^5 cells were harvested to perform the nuclease functional assay (see section 6. Materials and methods). Finally, at day 14 cells were harvested to conduct the second flow cytometry analysis and the integration analysis in the bulk population (see section 7.1. Materials and methods).

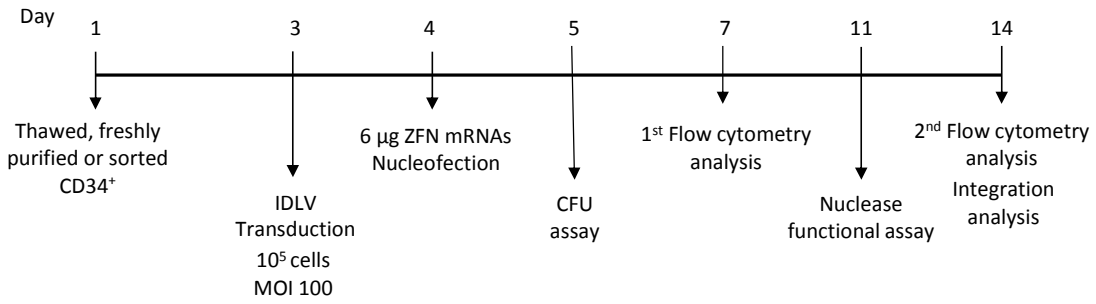


Figure 22: Schematic representation of the gene editing protocol in CD34+ cells.

5.2.1. Flow cytometry analysis in gene edited cells

1-5x10^4 cells were collected and resuspended in flow cytometry buffer (PBS containing 0.5% BSA and 0.05% azide) to stain during 30 minutes at 4°C with CD34 PE-Cy7, CD133 PE and CD90 APC (Table 6). Then, cells were washed and resuspended again in the flow cytometry buffer containing 1 µg/mL 4',6-diamidino-2-phenylindole (DAPI) as a viability marker.

Antibody	Clone	Supplier
CD34 PE-Cy7	4H11	eBioscience
CD133 PE	AC133	MACS
CD90 APC	5E10	BD

Table 6: List of antibodies used in flow cytometry for analysis of CD34+ gene targeted cells.

Analysis was performed in the LSR Fortessa cell analyser (BD/Becton, Dickinson and Company, New Jersey, USA). Off-line analysis was conducted with the FlowJo Software v7.6.5 (Tree star, Ashland, USA). In this analysis the percentage of GFP^+ cells was measured in the total population of cells and also in the different subpopulations of differentiated and progenitor cells. Gating strategy is shown in Figure 23: firstly, double positive $CD34^+$ $CD133^+$ cells were selected and, inside the double positive cells the percentage of CD90 positive and negative cells were selected. The most primitive compartment corresponds to the positive cells to the three markers ($CD34^+$ $CD133^+$ $CD90^+$), followed by early progenitors ($CD34^+$ $CD133^+$ $CD90^-$), committed progenitors ($CD34^+$ $CD133^-$ $CD90^-$) and finally, differentiated cells ($CD34^-$)[113].

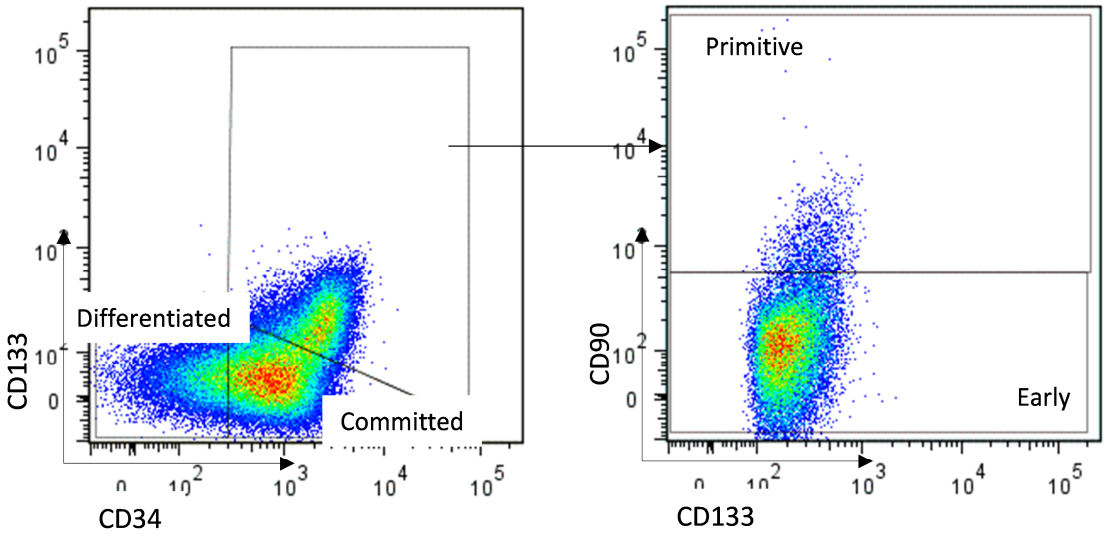


Figure 23: Gating strategy for the analysis of the different subpopulations of hematopoietic progenitor cells after gene editing.

5.2.2. Clonogenic assays

900 HD or 10,000 FA-A $CD34^+$ cells depending on the experiment were resuspended in 3 mL of methylcellulose. Each mL of the triplicate was seeded in a M35 plate and cells were incubated as shown in Table 3. Fourteen days later the number of colonies was counted using an inverted microscope (Olympus IX70 WH10x/22, objective 4x). CFUs-GMs (granulocyte-macrophage colonies) and BFU-Es (erythroid colonies) were identified.

Single colonies were picked to analyze the specific integration of the recombination cassette in the *AAVS1* locus.

6. Nuclease functional assay (Surveyor assay)

This assay determines the percentage of NHEJ efficiency by Surveyor nuclease assay in samples that were treated with the ZFNs. Firstly DNA of these cells was extracted using NucleoSpin® Tissue kit. Then, a PCR was performed to amplify the homology region of the *AAVS1* locus in which the ZFNs cut. The primers used in this PCR are NHEJ Fw (5' CTTCAGGACAGCATGTTTGC 3') and NHEJ Rv (5' ACAGGAGGENE TARGETINGGGGGTTAGAC 3'). The PCR was performed with 2 ng/µL of DNA, 1 µM of each primer, 1x Taq Extender Buffer, 0.12 U/µL Taq Extender PCR Additive (both from Agilent Technologies, Santa Clara, CA, USA), 0.12 U/µL AmpliTaq Gold DNA polymerase (Applied Biosystems) and 0.24 mM dNTPs (Invitrogen) in a final volume of 50 µL. The cycling conditions are the following: 95°C 10 minutes, 40 cycles of 95°C 45 seconds, 62°C 45 seconds and 72°C 45 seconds, and finally one cycle of 72°C 10 minutes. This PCR amplifies a 224 bp product that is then dehybridize and rehybridaze in order to obtain heteroduplexes that can be formed with a WT DNA strand, coming from a cell in which the ZFNs do not cut, and a mutant DNA strand, coming from a cell in which the ZFNs cut and this cut was repaired

by NHEJ, generating small insertions or deletions (INDELS). If an INDEL has occurred the heteroduplex presents a DNA hairpin that can be recognized by the Surveyor® nuclease (Surveyor® mutation detection kit (IDT, Coralville, Iowa, USA)) and cut, generating a band pattern that could be visualized on 10% TBE Gels 1.0 mm (Invitrogen) (Figure 24). Images from gels were analyzed in order to measure the cleavage fraction using Image J software (it measures the densitometry value of the different bands). The percentage of cleavage is determined by the next equation:

$$\%\ NHEJ = \frac{Cleavaged\ bands - (2 \times Background)}{(Cleavaged\ bands + Parental\ band) - (3 \times Background)} \times 100$$

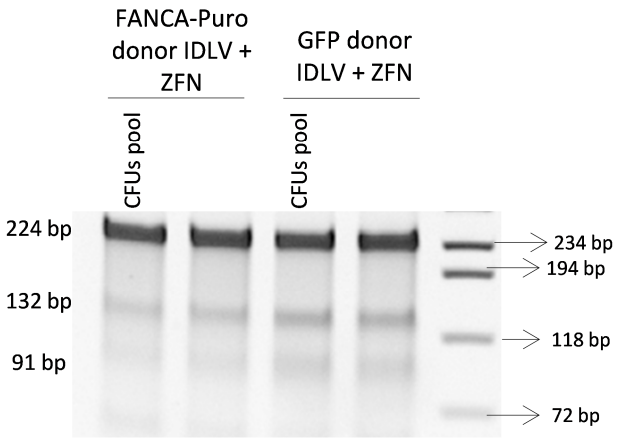


Figure 24: Representative image of Surveyor nuclease assay from one of the gene editing experiments in which a bulk population and also a pool of colony forming cells (CFCs) were assay. *The parental band of 224 bp is shown and also the two cleavage bands formed after nuclease treatment (132 and 91 bp).*

7. Analysis of the specific integration in the *AAVS1* locus

These assays allow us to confirm that the integration has occurred in the right place: the *AAVS1* locus.

7.1. PCRs

After gene editing assay, genomic DNA from bulk populations was extracted using NucleoSpin® Tissue kit. Single colonies were pelleted and resuspended in 10 µl of PBS. Their gDNA was extracted using 20 µl of lysis buffer (0.3 mM Tris HCl pH 7.5, 0.6 mM $CaCl_2$, 1,5 % Glycerol , 0.675 % Tween-20 and 0.3 mg/ml Proteinase K) and incubated at 65°C for 30 min, 90°C for 10 min and 4°C. After lysis, 30 µl of water was added, as previously described[302].

Two PCRs were conducted both for the 5' and 3' ends of the target integration site. A different pair of primers was used depending on the vector that was used as a donor template, but in all cases one of them hybridizes in the *AAVS1* locus, outside of the HA, and the other in the donor

(Figure 25). Using this approach amplification can only be detected when donor template is integrated in the *AAVS1* locus.

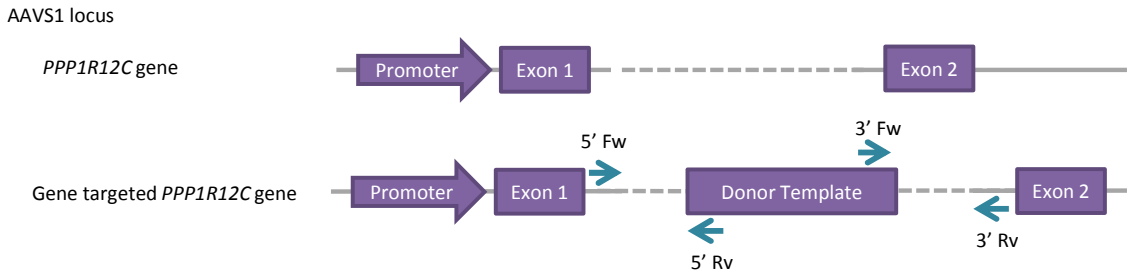


Figure 25: Representative figure of PCRs to confirm the targeted integration. *The arrows indicate the different primers used.*

PCR reactions were conducted using 100 ng of gDNA from bulk populations or 10 µL of gDNA from single colonies. PCRs from the 5' and 3' ends were performed with 1 µM of each primer (Table 7), 1x Taq Extender Buffer, 0.12 U/µL Taq Extender PCR Additive, 0.12 U/µL AmpliTaq Gold DNA polymerase and 0.24 mM dNTPs in a final volume of 25 µL. The conditions for the PCR were: 95°C 10 minutes, 40 cycles of 95°C 45 seconds, Tm 45 seconds and 72°C 2 minutes, and finally one cycle at 72°C 10 minutes.

Primer	Region		Sequence (5' to 3')	Tm (°C)	PCR product size (bp)
GFP donor IDLV	5'	Fw	AACTCTGCCCTCTAACGCTGC	59	1041
		Rv	ACGTGAAGAATGTGCGAGACCCAG	59	
	3'	Fw	AACGGGGATGCAGGGGAACG	59	997
		Rv	TTGCATCGCATTGTCTGAGTAGG	59	
FANCA-Puro donor IDLV	5'	Fw	AACTCTGCCCTCTAACGCTGC	59	1041
		Rv	ACGTGAAGAATGTGCGAGACCCAG	59	
	5'	Fw	ACTTCCCCTCTTCCGATGTT	54	1230
		Rv	GACGTGAAGAATGTGCGAGA	54	
GFP-FANCA donor IDLV	5'	Fw	AACTCTGCCCTCTAACGCTGC	59	1217
		Rv	TGGTGCAGATGAACTTCAGGG	59	
	3'	Fw	AACGGGGATGCAGGGGAACG	59	997
		Rv	TTGCATCGCATTGTCTGAGTAGG	59	
	3'	Fw	AAAGCTCGTCTTTTTCTGCTGCAGT	59	1332
		Rv	TTGCATCGCATTGTCTGAGTAGG	59	

Table 7: PCR primers used to amplify the targeted integration in the AAVS1 locus of the different donor constructs.

7.2. Vector copy number of donor construct in targeted cells

Q-PCR was used to determine the VCN in gene targeted cells and also in samples from different organs from mouse engraftment experiments using gDNA as template. gDNA was extracted with NucleoSpin® Tissue kit. This template was amplified using specific primers listed in Table 8. Q-PCR amplification was performed in a 7500 fast real-time PCR system using TaqMan and fluorescent probes that were acquired from Sigma and are listed in the Table 8.

Primer/Probe		Sequence (5' to 3')	Tm (°C)
FANCA	Fw	GCTCAAGGGTCAGGGCAAC	58
	Rv	TGTGAGAAGCTCTTTTTCGGG	60
	Probe	[FAM]-CGTCTTTTTCTGCTGCAGTTAATACCTCGGT-[BHQ1]	70
GFP	Fw	GTAAACGGCCACAAGTTCAGC	61
	Rv	TGGTGCAGATGAACTTCAGGG	61
	Probe	[FAM]-CTTGCCGTAGGTGGC-[BHQ1]	43.9
Psi	Fw	TCCCCCGCTTAATACTGACG	59
	Rv	CAGGACTCGGCTTGCTGAAG	63
	Probe	[FAM]-CGCACGGCAAGAGGCGAGG-[BHQ1]	75.7
hAlb	Fw	GCTGTCATCTCTTGTGGGCTG	64
	Rv	ACTCATGGGAGCTGCTGGTTC	63
	Probe	[TxRd]-CCTGTCATGCCCACACAAATCTCTCC-[BHQ2]	73.9

Table 8: Q-PCR primers and probes used to determine VCN by Q-PCR.

8. Functional assays in gene edited cells

These assays were performed in order to determine the functionality of the *FANCA* gene in cells targeted by gene editing.

8.1. Nuclear foci formation

Cells were seeded in chamber slides (Nunc, Sigma) pre-coated with retronectin® (20 µg/cm^2 during 24 hours at 4°C (Takara bio inc., Japan)) at confluence. 24 hours later cells were treated with 40 nM MMC (Sigma) for 16 hours to induce DNA damage and foci formation. Then, cells were washed with PBS 1X and fixed with 3.7% paraformaldehyde (Sigma) 15 minutes at room temperature. Cells were washed 3 times with TBS 1X (20 mM Tris-HCl + 140 mM NaCl) and permeabilized using 0.5% Triton X100 (Sigma) during 10 minutes, washed again and blocked with blocking solution (0.1% Nonidet-P40 (Sigma) + 10% Hyclone) during 45 minutes. Cells were incubated over night at 4°C in a humidity chamber with the primary antibody against human FANCD2 (ab2187-50, Abcam, Cambridge, UK) or human RAD51 (ab46981-100, Abcam) diluted 1:250 in blocking solution. Cells were washed 3 times with TBS 1X and then incubated during one hour with the secondary antibody IgG against rabbit, conjugated with TexasRed (711-075-152, Jackson ImmunoResearch, West Grove, PA, USA) 1:500 dilution and DAPI (Roche) 1:1,000 dilution in blocking buffer. Finally cells were washed 3 times in TBS 1X and slides were mounted with Moviol.

Samples were visualized with a fluorescence microscope Axioplan 2 imaging (Carl Zeiss, Jena, Germany) with the objective 0.17 mm and 100x/1.45 magnification. Two hundred cells were counted and cells with more than 10 foci were analyzed in each sample. Images were captured with AxioCam MRm (Carl Zeiss) and processed with AxioVision 4.6.3 (Carl Zeiss) and Corel Photo-Paint 11 (Corel, Ottawa, Ontario, Canada).

8.2. hFANCA expression by western blot

Western blot analysis was performed to determine the expression of the human FANCA protein in LCLs samples transduced with the FANCA-Puro donor vector. Protein was extracted

from 10^7 cells per condition using: 20 mM Hepes-NaOH, 12.5% glycerol, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM PPiNa, 20 mM FNa, 1 mM orthovanadate and 1x protease inhibitors (Mini-complete, Roche). Samples diluted in the extraction buffer were frozen in liquid nitrogen and warmed in a thermoblock at 37°C five times, then 1% of Nonidet-P40 was added and samples were shaken at 4°C during 15 minutes. Finally, samples were centrifuged and the supernatant, in which the proteins are, was collected. Protein quantification was carried out using Bradford method (BioRad, Hercules, CA, USA) in a spectrophotometer.

Fifty µg of each sample was used in NUPAGE 4-12% Bis-Tris poliacrilamyde gels (Novex/Life Technologies/Thermo Fisher Scientific, Waltham, USA) in redox and denaturing conditions with MOPS buffer (NuPAGE, Novex). Electrophoresis was run for 7 hours at 4°C and 100 V. Then, samples were transferred to PDVF membrane during 2 hours at 90 V and 4°C in transfer buffer (NuPAGE). After the transference blocking of unspecific binding sites was performed with 5% powder milk diluted in TBS-T buffer (TBS 1X + 0.1% Tween) in agitation at RT during 1 hour. Then the membrane was stained with anti-hFANCA antibody produced in rabbit (ab5063, Abcam) (1:1,000 dilution in 0.5% milk TBS) or anti-β-actin antibody produced in mouse (ab6276, Abcam) (1:4,000 dilution in 0.5% milk TBS) over night at 4°C. The membrane was washed with TBS-T buffer three times and then was incubated with 1:5000 dilution in 0.5% milk TBS of the secondary antibody anti-rabbit (pAb to Rb IgG-HRP, ab6721-1, Abcam) anti-mouse (Shp pAb to Ms IgG-HRT ab6808, Abcam) during 1 hour at RT. Finally, the membrane was washed three times with TBS-T.

Membranes were developed by chemiluminiscence signal detection (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific, Pierce, USA) through photography film (Amersham, UK) exposure in an automatic developer (Curix60, AGFA, Mortsel, Belgium).

8.3. MMC resistance

Gene targeted LCLs were evaluated in order to see the reversion of the phenotype. After the gene targeting protocol (see section 5.1. Materials and methods) $5x10^4$ cells were seeded in 24 well plates in a final volume of 1 mL and increasing concentrations of MMC were used (from 0 to 1,000 nM). Cell viability was determined 5 to 15 days after by flow cytometry using DAPI vital marker in a LSRFortessa cell analyser and the off-line analysis was performed with the FlowJo Software v7.6.5.

8.4. Transplant in immunodeficient mouse

NOD.Cg-*Prkdcscid*Tg (hSCF/hGM-CSF/hIL3) (NSG-SGM3) mice were used in these experiments. Mice were sublethally irradiated with 1.5 Gy the day before transplant. $4.5x10^5$ $CD34^+$ cells in which gene editing had been conducted (see section 5.2. Materials and methods) were transplanted through the tail vein at day 6 of the experiment (Figure 26). 30 and 60 days after transplant bone marrow (BM) aspiration was performed to analyze the human hematopoietic engraftment and the percentage of gene edited cells by flow cytometry analysis of hCD45 APC and hCD34 PE. 90 days post-transplant (dpt) mice were sacrificed and peripheral blood (PB), spleen and BM were collected. These organs were processed and stained to analyze human reconstitution, the percentage of gene targeted cells and also multilineage engraftment by flow cytometry using different antibodies (Table 9). The percentage of gene targeted cells and

the multilinage engraftment of the different subpopulations was analyzed in the hCD45^{+} fraction as it is shown in Figure 27.

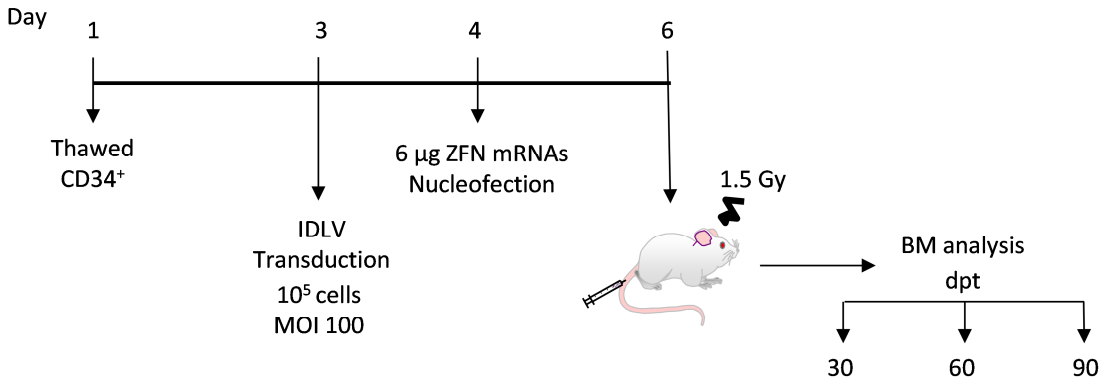


Figure 26: Protocol used to analyze the engraftment capacity of gene edited cells in immunodeficient mice. *BM: bone marrow; dpt: days post-transplant.*

Antibody	Clone	Supplier
hCD34 APC	581	BD
hCD45 APC-Cy7	HI30	BioLegend
hCD33 PE-Cy7	P67.6	BD
hCD3 APC	UCHT1	BioLegend
hCD19 PE-Cy7	SL25C1	eBioscience
hCD45 APC	2D1	eBioscience
hCD34 PE	8G12	BD

Table 9: Antibodies used to determine the multilineage human engraftment in immunodeficient mice transplanted with gene targeted cells.

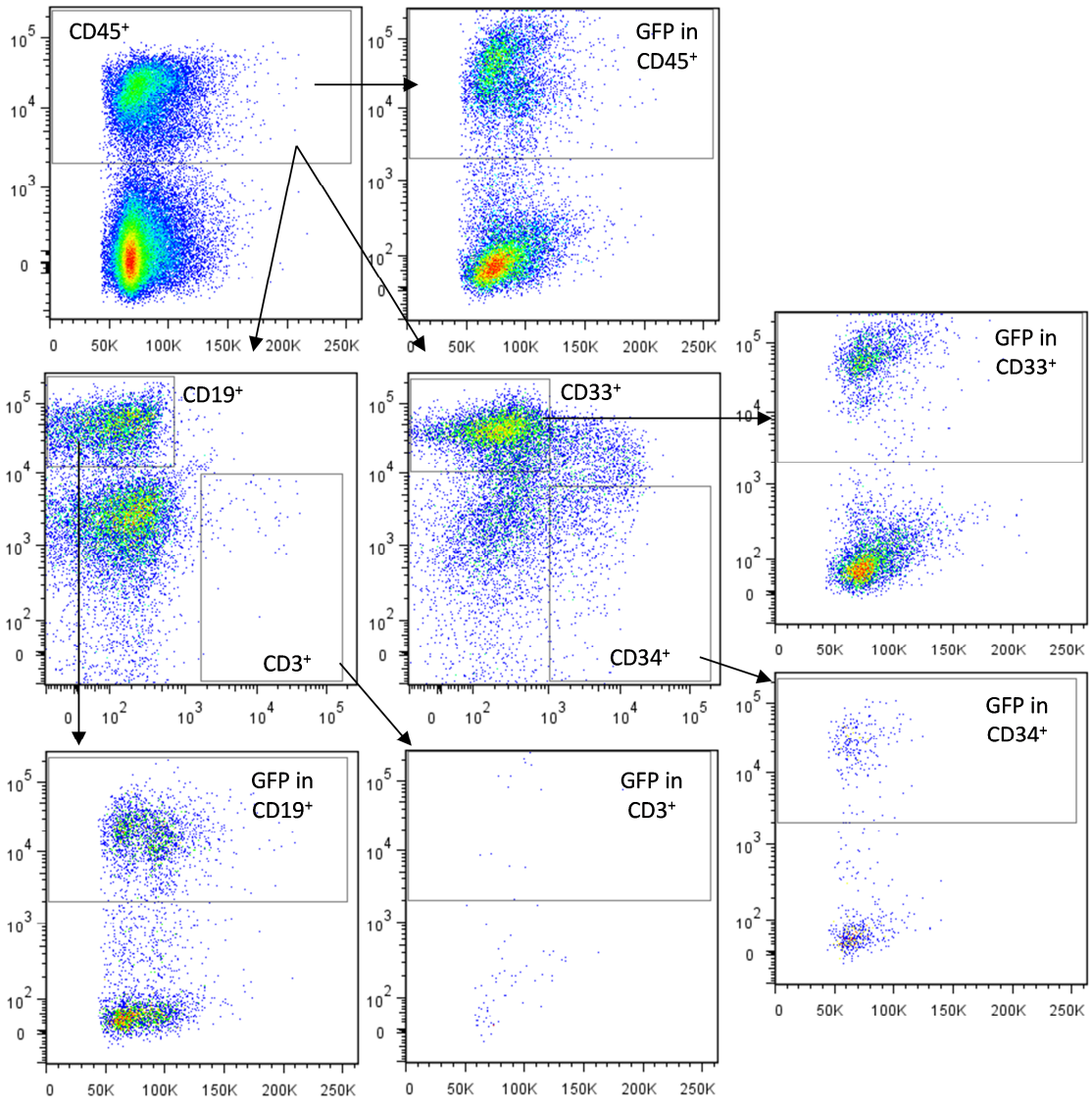


Figure 27: Multilineage analysis in BM and spleen of NSG-SGM3 mice. *hCD45$^+$ cells were selected and hCD3$^+$, hCD19$^+$, hCD33$^+$, hCD34$^+$ and GFP were gated in this population. Finally, GFP positive cells were also analyzed in the different subpopulations.*

Human CD45$^+$ cells from BM of these animals were purified using immunomagnetic selection beads anti-hCD45$^+$ (MACS). 10^5 cells were used in clonogenic assays (see section 5.2.2. Materials and methods).

9. Gene targeting improvement

9.1. Improvement of homology directed repair efficiency by PARI inhibition

9.1.1. Improvement of gene editing in lymphoblastic cell lines

$2x10^5$ HD or FA LCLs were nucleofected with 25 pmol of a mix of three different PARI iRNAs (5' GGCAGAGAUACUGUGAAAGCCUUAU 3'; 5' CCAUCAUCAUGGAACGUCUAUUCUU 3'; 5' CAGUGAUAAAGAUGCAACUGAUUAA 3') (Stealth, Invitrogen) using SF Cell Line 4D-Nucleofector® X Kit. As iRNA control we used Block-iT™ (Invitrogen).

To verify that PARI inhibition correlates with inhibition of HR, RAD51 foci were stained as described in section 8.1 Materials and methods.

Gene editing experiments were conducted 24 hours after the nucleofection using MOI 50 of GFP donor IDLV and 1.5 µg of pVAX ZFN pair as mRNA as described in section 5.1.2. Materials and methods.

9.1.1.1. Analysis of PARI inhibition

PARI interference was evaluated by Q-PCR two days after transfection. Cells were collected; RNA was extracted using High Pure RNA isolation kit (Roche) and RT-PCR was performed using RETROscript® First Strand Synthesis (Ambion). Finally, Q-PCR was conducted using SYBRGreen FastStart Universal SYBR Green Master PCR master mix (Roche) in Rotor Gene 6000 (Corbet, Australia) with specific primers (Table 10).

Primer		Sequence (5'to 3')	Tm (°C)
PARI	Fw	TGTAGCTCTTAGCACCACTGACA	63
	Rv	CCATGGTTTATGGCATGAGA	56
β-Actin	Fw	ATTGGCAATGAGCGGTTCC	54.7
	Rv	CACAGGACTCCATGCCCA	51.9

Table 10: Primers to measure the inhibition of PARI by Q-PCR.

9.1.1.2. Chromosomal stability analysis

The possible increment in genomic instability produced by PARI inhibition was studied in metaphases of interfered cells. Briefly, interfered cells were treated or not during 72 hours with 0.1 µg/mL of diepoxybutane (DEB) (Sigma) to induce DNA damage. Then 0.1 µg/mL of colchemide (Gibco) was added to the culture media and cells were incubated for 4 hours to induce a block of cell division in metaphase. At that time cells were treated with 0.56% of KCl (0.075M) to provoke an osmotic shock. Finally, cells were fixed with methanol:acetic acid (4:1) solution and dropped onto clean slides. Metaphases were stained with 1% Giemsa stain, and in a total of 25-40 metaphases per sample, chromosomal aberrations were counted.

9.1.2. Improvement of gene editing in fibroblasts

Human fibroblasts immortalized with hTert were interfered with a single PARI iRNA (5' GGCAGAGAUACUGUGAAAGCCUUAU 3') (Stealth, Invitrogen) in order to increase the efficiency of HDR. As iRNA control Block-iT™ was used. Two rounds of transfection with 25 pmol of iRNA

were performed every 24 hours using Lipofectamine 2000 (Invitrogen), according to manufacturer instructions. PARI interference was evaluated by Q-PCR two days after the last transfection as it is described in section 9.1.1.1. Materials and methods.

One day after the last transfection the gene targeting was performed, transduction was done with MOI 5 of GFP donor IDLV. The day after $5x10^5$ cells were nucleofected with 3 µg of total pVAX ZFNs as mRNA (1.5 µg of each ZFN) using Amaxa® Human Dermal Fibroblast Nucleofector® Kit (NHDF) (Lonza). Finally the percentage of GFP positive cells was measured by flow cytometry 3 and 10 days after nucleofection.

9.2. Gold nanoparticles to improve transduction efficiency

$CD34^+$ cells purified from UCB were prestimulated during 48 hours as previously described. The integrative competent and integrase-defective version of the GFP donor lentivirus were incubated one hour at room temperature with 4-8 nM of gold nanoparticles (AuNP) coated with positive charge and conjugated with TAMRA fluorescence protein. AuNPs were gently provided by Dr. Jesús M. de la Fuente (Instituto de nanociencia de Aragón, Zaragoza, Spain). Then $5x10^4$ $CD34^+$ prestimulated cells were transduced with different MOIs of these lentiviral vectors, previously incubated or not with the AuNPs. One day after transduction cells were washed.

Cell viability was measured by DAPI exclusion in flow cytometry analysis performed one and two days after treatment, and also by trypan blue under microscope in which the proliferation rate was calculated by the number of living cells counted divided by the number of cells seeded the day before.

Transduction efficiency was measured by flow cytometry analysis of TAMRA and GFP expression. In the case of cells transduced with IDLV vectors these parameters were measured 1 and 2 days post transduction and in the case of the ICLV the transduction efficiency was measured 2 and 12 days post-transduction.

9.2.1. Evaluation of human transduced cells engraftment in NSG mice

One day after transduction, 250,000 cells that had been transduced with the GFP donor ICLV, either alone or with the AuNPs, were transplanted through the tail vein in NOD.Cg-*Prkdcscid IL2rgtm1Wjl*/SzJ (NSG) mice that had been sublethally irradiated the day before transplant with 1.5 Gy. Analysis of human engraftment was conducted as previously described in the section 8.4. Materials and methods. The thymus was also collected at the end point of the experiment, processed and analyzed using the antibodies listed in Table 11. In this organ the analysis of $CD4^+$ and $CD8^+$ subpopulations is done inside the $CD3^+$ subpopulation as it is shown in Figure 28.

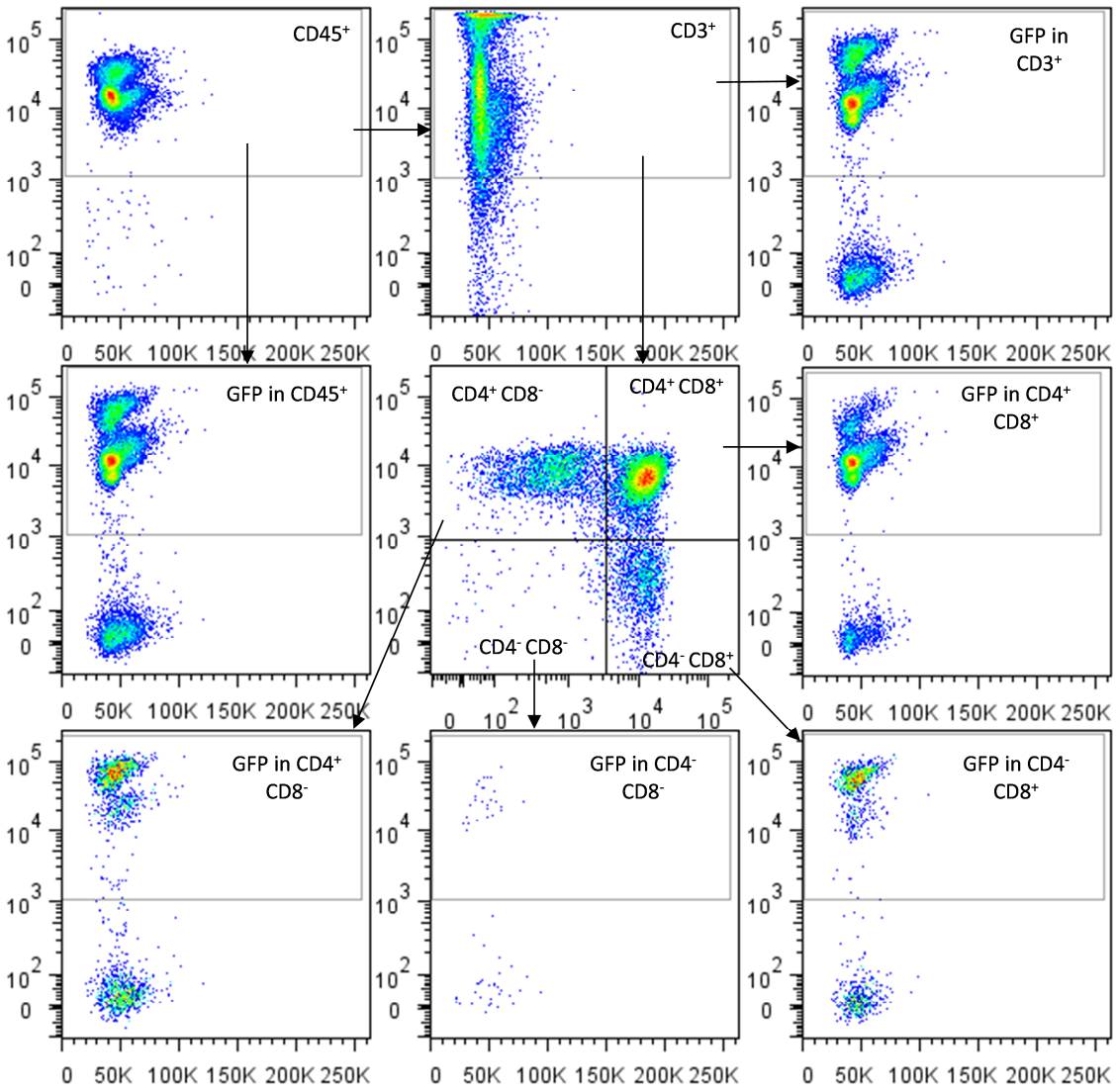


Figure 28: Multilineage analysis in the thymus of NSG mice. *$hCD3^+$ cells were analyzed in the $hCD45^+$ cells. GFP positive cells were gated in h $CD45^+$ and $hCD3^+$ populations. $hCD4^-$ $hCD8^-$, $hCD4^+$ $hCD8^+$, $hCD4^+$ $hCD8^-$ and $hCD4^-$ $hCD8^+$ were gated inside $hCD3^+$ population. Finally, the GFP population inside all this subpopulations was analyzed.*

Antibody	Clone	Supplier
hCD3 PE-Cy7	UCHT1	BioLegend
hCD4 APC	MT466	Miltenyi Biotech
hCD8 PE	B9.11	Immunotech

Table 11: List of antibodies used to determine multilineage engraftment in the thymus of immunodeficient mice transplanted with gene targeted cells.

Secondary transplants were conducted using 80% of bone marrow from primary recipients. Animals were transplanted one to one into sublethally irradiated recipients (1.5 Gy dose the day before).

Analysis of these secondary recipients was performed as previously described in section 8.4. Materials and methods.

10. Statistical analysis

The statistical analysis was performed using GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego, USA). For the analyses of experiments in which $n<5$, a non-parametric two-tailed Mann-Whitney test was performed when two variables were compared or Kruskal-Wallis with Dunn's multiple comparison test when more than two variables were compared. In the experiments in which $n\geq5$, a Kolmogorov-Smirnov test was done to test the normal distribution of the samples. If samples showed a normal distribution, a parametric two-tailed paired T-test was performed when two variables were compared or an ANOVA with Bonferroni's multiple comparison test when more than two variables were compared. If samples did not follow normal distribution, the previously mentioned non-parametric tests were used. The significances are indicated as $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

RESULTS

1. Gene targeting with a reporter and therapeutic *FANCA* genes in healthy donor hematopoietic cells

1.1. Analysis of gene targeting efficiency in lymphoblastic cell lines

1.1.1. Evaluation of ZFN efficiency in lymphoblastic cell lines

The activity of two different *in vitro* synthetized mRNA constructs, TEV and pVAX, carrying the left and right ZFNs targeting the *AAVS1* locus was tested in a healthy donor (HD) lymphoblastic cell line (LCL) (C4) (Figure 29). The principal difference between both constructs is that TEV presents two regulatory regions: 5' UTR from the TEV and 3' UTR from human *β-globin*; whereas the pVAX construct does not present any regulatory region. The indicated doses represent the total amount of left and right ZFNs for each construct. For example, the 3 µg dose means that cells were nucleofected with 1.5 µg of the right ZFN and 1.5 µg of the left ZFN. Nucleofected cells were harvested five days later to measure the activity of ZFNs using the Surveyor assay. In all instances, the efficacy of pVAX ZFN pair was higher, with more than 50% of insertions/deletions (INDELs) representative of NHEJ repair.

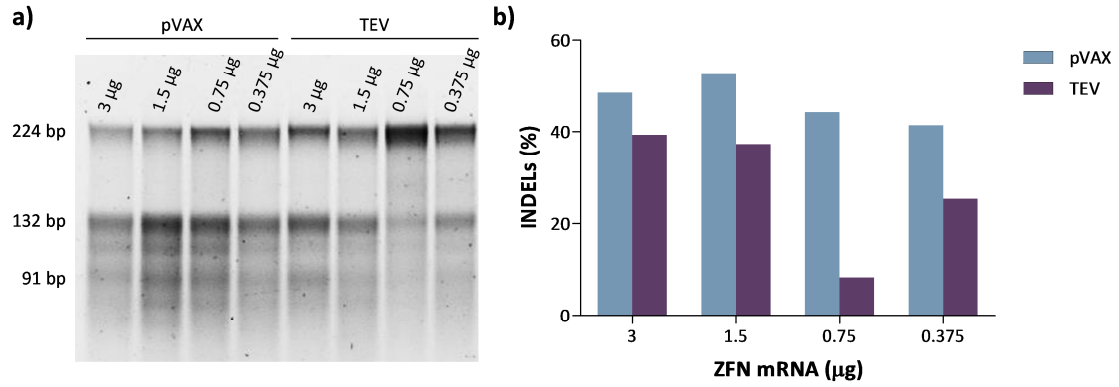


Figure 29: Surveyor assay of C4 LCL with different doses of pVAX and TEV ZFNs. *a) Electrophoresis gel corresponding to the Surveyor assay. The parental band (224 bp) and two processed bands (132 and 91 pb) are highlighted. b) Quantification of the percentage of INDELs calculated by measuring the ratio of the processed and parental bands using the ImageJ software.*

Thereafter, a gene targeting experiment using a GFP donor packaged in an IDLV was performed in this cell line to verify the efficiency of HDR using different doses of each ZFN pair. Consistent with the results of the Surveyor assay, higher percentages of gene targeting (measured as the percentage of GFP^+ cells at day 12 of culture) were achieved when the pVAX ZFN pair was used (Figure 30 b). Only with the highest dose of the pVAX ZFN pair, gene targeting was low, most probably due to the toxicity associated to this dose (Figure 30 a).

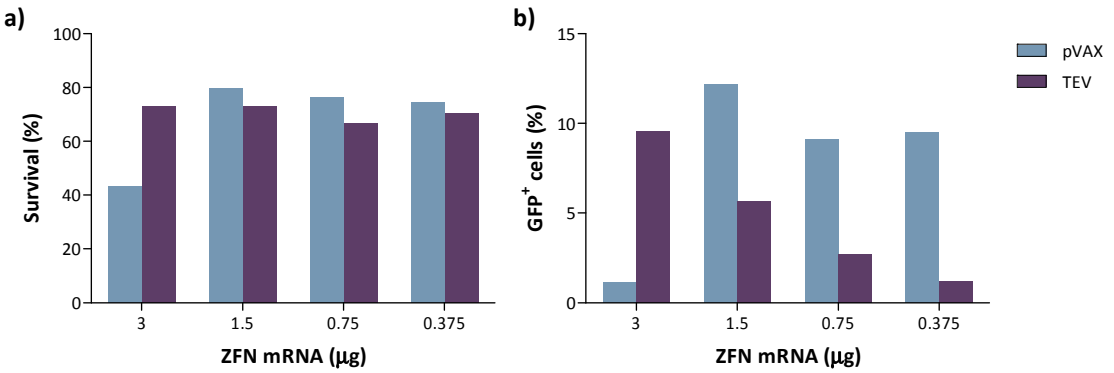


Figure 30: Efficiency of different ZFN pairs in gene targeting experiments using a healthy donor LCL. *Cells were nucleofected with different amounts of each ZFN pair. The viability and percentage of GFP$^+$ cells was measured. a) Percentage of alive cells 3 days after nucleofection measured by DAPI exclusion in flow cytometry analysis. b) Percentage of GFP positive cells evaluated by flow cytometry analysis in the different experimental conditions at day 12 of culture.*

The integration in the *AAVS1* locus was analyzed by PCR in the samples with the highest percentage of GFP$^+$ cells (3 µg in the case of TEV pair and 1.5 µg in the case of pVAX pair). The PCR was conducted in the bulk population using primers that hybridize in the endogenous locus of the *AAVS1*, outside of the homology arms (HAs), and inside the donor construct, both for the 5' and the 3' integration ends (Figure 31 a). As it was expected there was only amplification of the specific bands in the samples that were treated with the donor and the ZFNs (Figure 31 b). Although in the sample targeted with pVAX ZFNs only the band of the 3' end was amplified, this could be due to the PCR reaction as in the rest of the experiments we were able to amplify this band. Taking into account the higher efficiency of gene targeting using pVAX construct, we selected this ZFN pair as the construct for the rest of the experiments conducted using LCLs.

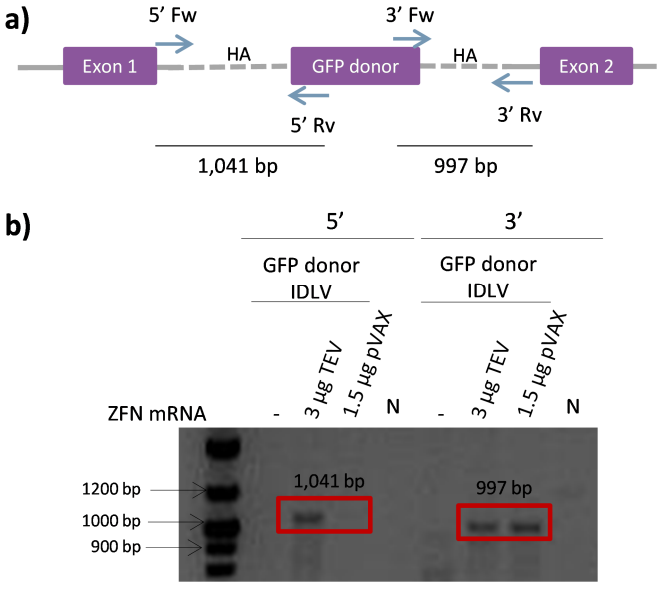


Figure 31: Representative PCR of integration in the AAVS1 locus using GFP donor IDLV in a HD LCL*. a) Gene targeting of the GFP donor IDLV into the AAVS1 locus. The arrows represent different primers used*

to detect the specific integration. b) Integration was confirmed in C4 LCL after gene editing in both the 5' and 3' ends (except for the 5' end of the pVAX sample). HA: homology arm. Fw: forward. Rv: reverse. N: negative control.

The fact that mRNA transfection mediated a high and transient expression of the ZFNs, reducing the risk of random insertion, led us to select the pVAX mRNA to perform the rest of the experiments.

1.1.2. Selection of the donor delivery system: plasmid DNA versus IDLV

In order to test the efficacy of the donor construct delivered as plasmid DNA in HD LCLs (C4), we nucleofected these cells with our GFP donor plasmid and the ZFN mRNA. When the donor was delivered as plasmid DNA, flow cytometry analysis conducted at day 12 of culture revealed that 2.99±0.22% of the population was GFP^+. Very small population of cells (0.18±0.01%) were GFP^+ when only the donor plasmid was used, suggesting that random integration was occurring in a low proportion of cells (Figure 32 a). The mean percentage of INDELs measured by Surveyor assay in samples nucleofected with the plasmid GFP donor and the ZFNs was 51.43±1.97%.

The PCR integration analysis of cells nucleofected with the donor plasmid confirmed the specific integration in the *AAVS1* locus only when the samples were nucleofected with the donor and the ZFNs (see PCRs for both integration ends. Figure 32 b).

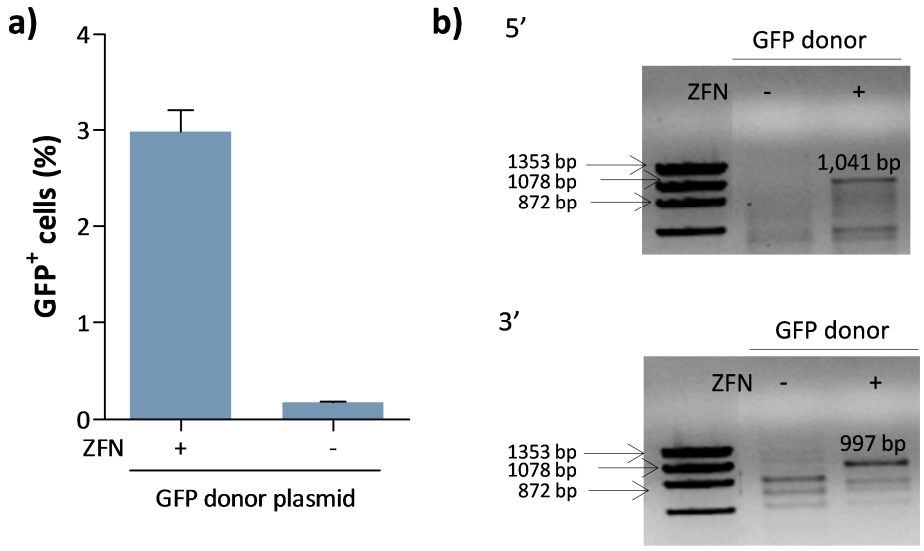


Figure 32: Gene targeted HD LCL (C4) using the donor as plasmid DNA. *a) The percentage of GFP^+ cells was analyzed by flow cytometry at day 12 of cell culture. Bars represent the mean and standard deviation (SD). n=2. b) PCR of the integration in the AAVS1 locus for the 5' and 3' ends.*

Trying to improve the gene editing efficiency and also to minimize the random integration, we also tested the donor as an IDLV. Several independent experiments were performed in LCLs from two HDs (CP1 (n=5) and C4 (n=6)). The percentage of GFP^+ cells was measured at day 12 of cell culture after transduction only with the GFP donor IDLV, or also nucleofected with the ZFNs mRNA. The combination of the GFP donor IDLV and the ZFNs mRNA led to higher percentages of GFP^+ cells than when the donor plasmid was used, reaching a 5.13% of GFP^+

cells in CP1 and 9.22% in C4 LCLs. The use of GFP donor IDLV alone also generated a low random integration of the donor cassette, measured by the presence of GFP^+ cells (0.1% in CP1 and 0.24% in C4 LCLs) (Figure 33). Moreover, the efficiency of the gene targeting process was cell line dependent, as the percentage of CP1 GFP^+ cells (5.13±2.18%) were significantly lower than the percentage of C4 GFP^+ cells (9.22±4.1%).

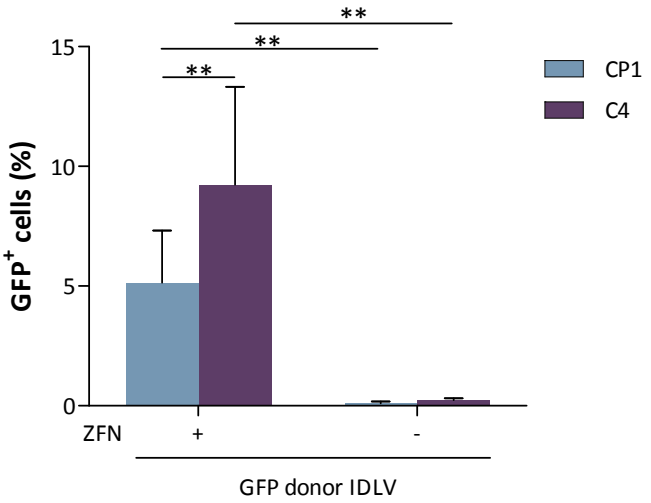


Figure 33: Gene targeting of GFP in HD LCLs using IDLVs. *Two different HD LCLs were targeted using a GFP donor IDLV. The percentage of GFP^+ cells was evaluated by flow cytometry at day 12 of culture. Bars represent the mean with SD. n=5 in CP1 and n=6 in C4. The significance of differences between groups is calculated using T-test, p<0.01 (**).*

To verify that the integration took place in the *AAVS1* locus, a PCR in the 5' and 3' ends of nucleofected cells was performed. PCR products confirmed the integration in the specific locus in both cell lines (Figure 34).

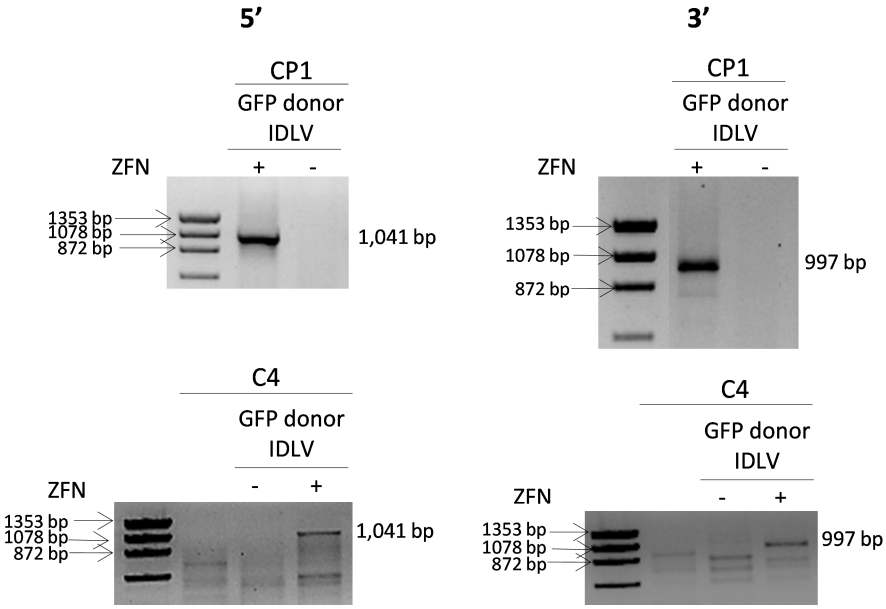


Figure 34: Representative PCR of the integration of the GFP transgene integrated in the AAVS1 locus in HD LCLs. *Gene targeting was conducted in CP1 and C4 LCLs and the PCR for the targeted integration of*

the GFP donor IDLV in the 5' and 3' ends was conducted in the bulk population at day 12 of cell culture. Top: CP1 PCR for AAVS1 specific integration. Bottom: C4 PCR for AAVS1 specific integration.

As PCR is the main test to verify if the integration is site-specific, we analyzed the detection limit of this assay using the GFP donor IDLV. Serial dilutions of the DNA from gene targeted C4 LCL, in which the percentage of gene targeting measured by flow cytometry was 11%, were analyzed by PCR. As shown in Figure 35, a detection limit of 2.25% gene targeting (GFP^+ cells) was observed in this PCR assay, indicating that levels of gene targeting lower than this percentage will not be detectable in bulk populations by this technique.

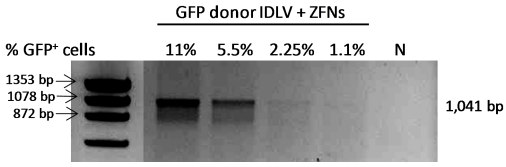


Figure 35: Detection limit analysis of PCRs used to evaluate the specific integration of GFP donor IDLV in the AAVS1 locus. *C4 LCL was gene targeted using GFP donor IDLV and the percentage of gene targeting was evaluated by flow cytometry at day 12 of culture. Cells were harvested and DNA from the bulk population was extracted. Serial dilutions of this DNA were used to conduct the PCR from the 5' end to determine the limit of detection of the assay. N: negative control.*

1.1.3. Gene targeting efficiency of a GFP-FANCA donor IDLV in LCLs

Aiming to evaluate the integration efficacy of the therapeutic GFP-FANCA donor, previously used to correct the phenotype of fibroblasts from FA-A patients[75], C4 LCLs were targeted using this donor IDLV vector and two different ZFN constructs (pVAX and TEV). Once integrated in the *AAVS1* locus the GFP expression is driven by the endogenous promoter of the *AAVS1* locus, while FANCA expression is driven by PGK promoter. Surprisingly, when these cells were analyzed by flow cytometry at day 12 of culture, GFP^+ cells could not be easily detected (Figure 36). Nevertheless, cells targeted with PGK-GFP construct showed evident GFP expression, suggesting a very weak activity of the *AAVS1* endogenous promoter in LCLs, in comparison with the exogenous PGK promoter used in the GFP donor vector.

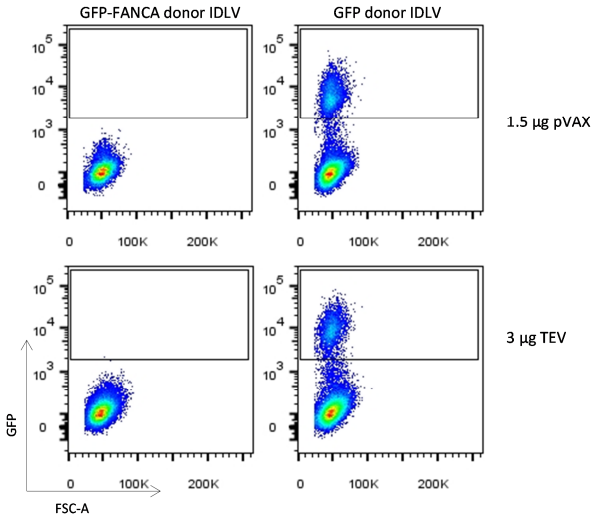


Figure 36: Representative dot plot of C4 LCLs gene edited with GFP-FANCA or GFP donor IDLVs using the pVAX or TEV ZFN constructs. *HD LCLs were targeted using GFP-FANCA and GFP donor IDLVs*

respectively with the two different ZFN pairs. The percentage of GFP positive cells was evaluated at day 12 of culture by flow cytometry. Images were obtained using FlowJo Software.

Despite the difficulty to detect GFP$^+$ cells when the GFP-FANCA donor IDLV was used, we investigated the integration of this construct in the *AAVS1* site (Figure 37 a). All the PCRs showed the integration of the donor cassette in the *AAVS1* site (Figure 37 b), indicating that the almost complete absence of GFP expression should be a consequence of the low expression driven by the *AAVS1* promoter.

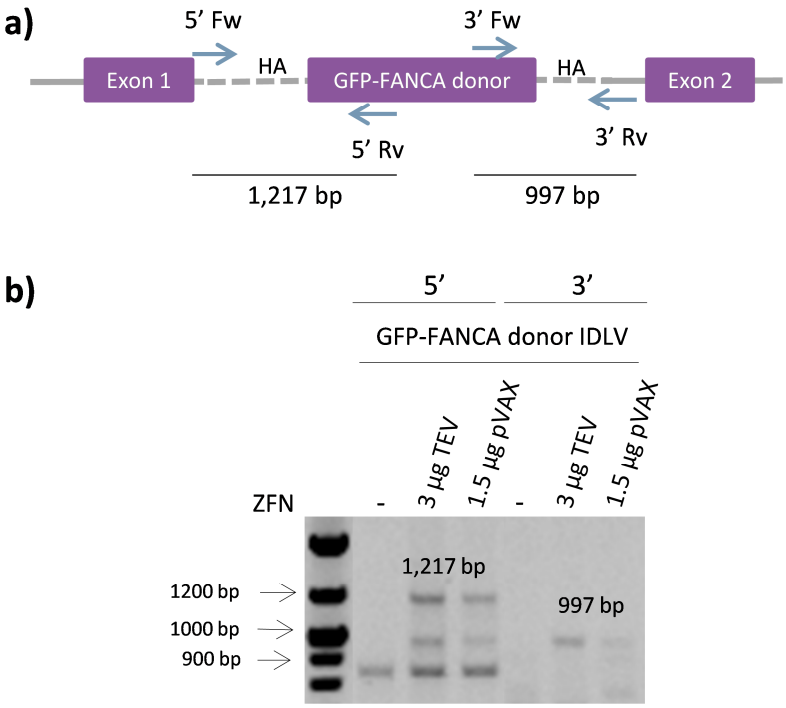


Figure 37: Representative PCR of the integration in the AAVS1 locus of GFP-FANCA donor used as IDLV. *a) Schematic representation of the GFP-FANCA donor specifically integrated in the AAVS1 locus. The arrows represent the primers used to evaluate if integration was site-specific. b) C4 LCLs were targeted and the PCR for the targeted integration of the GFP-FANCA donor IDLV in the 5' and 3' ends was conducted in the bulk population.*

1.2. Gene targeting in primary human hematopoietic progenitor cells

1.2.1. Optimization of the gene editing conditions in umbilical cord blood CD34$^+$ cells

1.2.1.1. Evaluation of the activity of two ZFNs in umbilical cord blood CD34$^+$ cells

Once the conditions and reagents were tested in cell lines from HD, we moved to our main target population, the human hematopoietic stem and progenitors cells (HSPCs), characterized by the expression of the CD34$^+$ antigen in their membrane. Firstly, and as we did with LCLs, two different doses of both ZFNs pair constructs, used as mRNA, were evaluated in CD34$^+$ cells from umbilical cord blood (UCB) using a multiplicity of infection (MOI) of 100 LVs/cell of the GFP donor IDLV. Analysis of GFP positive cells at day 14 of cell culture showed gene targeting only when the pVAX construct was used (Figure 38 a). This study also demonstrated that the

higher concentration of the ZFNs used in the experiments allowed us to obtain levels of GFP$^+$ cells of 15.7%. The surveyor analysis of these cells confirmed that only the pVAX pair generated INDELs in human progenitors (Figure 38 b). Higher doses of pVAX ZFNs (up to 12 µg) were also tested in these cells but low percentage of GFP$^+$ cells was observed, probably due to an increment in cell toxicity (Figure 38 c).

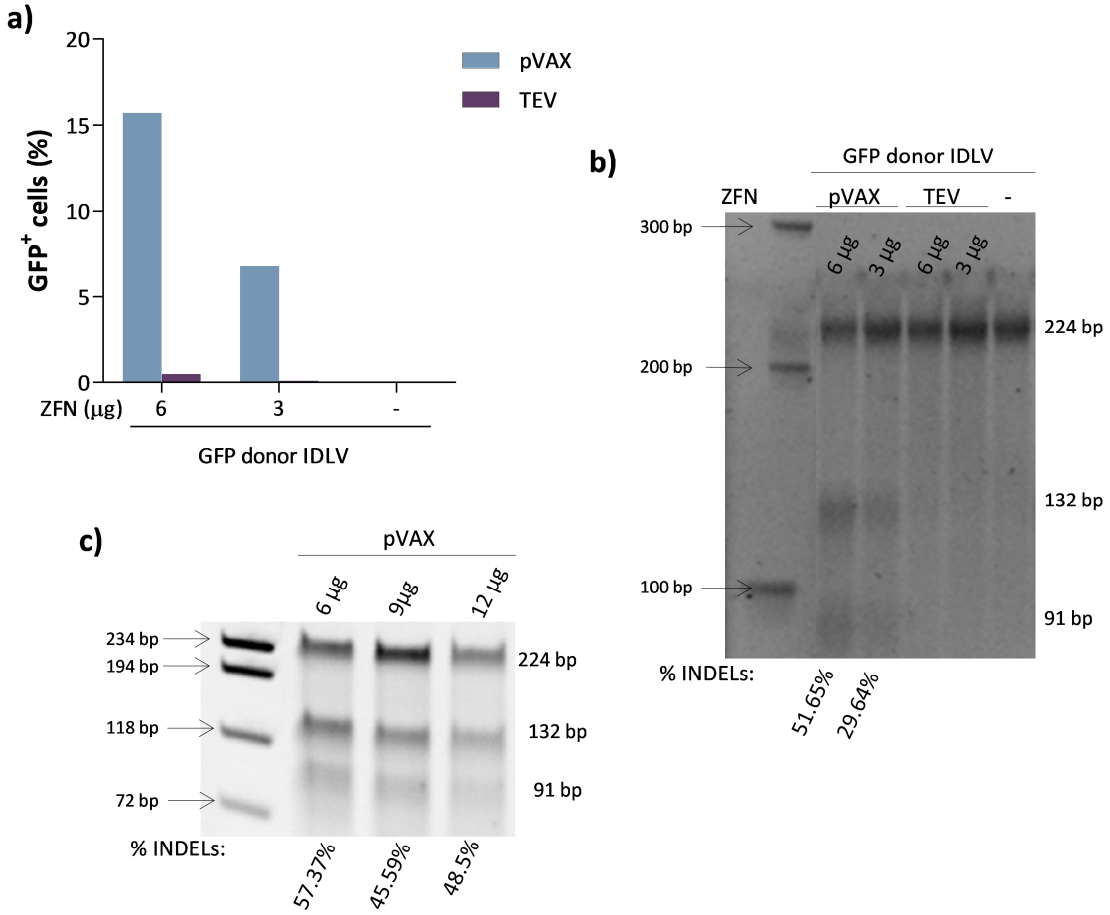


Figure 38: Evaluation of the efficiency of ZFNs mRNA to target CD34$^+$ cells. *a) Percentage of GFP$^+$ CD34$^+$ cells measured by flow cytometry at day 14 of cell culture, using two different doses of each ZFN pair: pVAX and TEV. b) Surveyor assay in gene targeting experiment showed in panel a. Size of the Surveyor bands and the percentage of INDELs induced by selected ZFNs are also indicated. c) Surveyor assay evaluating increased doses of pVAX ZFNs. Surveyor bands and percentage of INDELs are shown.*

To confirm that the GFP expression observed in CD34$^+$ cells was due to the specific integration in the *AAVS1* locus, integration analyses were performed by PCR in cultured cells transduced with the GFP donor IDLV and nucleofected with both types of ZFN pairs. As shown in Figure 39 b; gene targeting in the *AAVS1* locus only occurred when pVAX ZFNs were used. Based on these results we also selected the pVAX ZFN pair as the best alternative for our gene targeting experiments in CD34$^+$ cells.

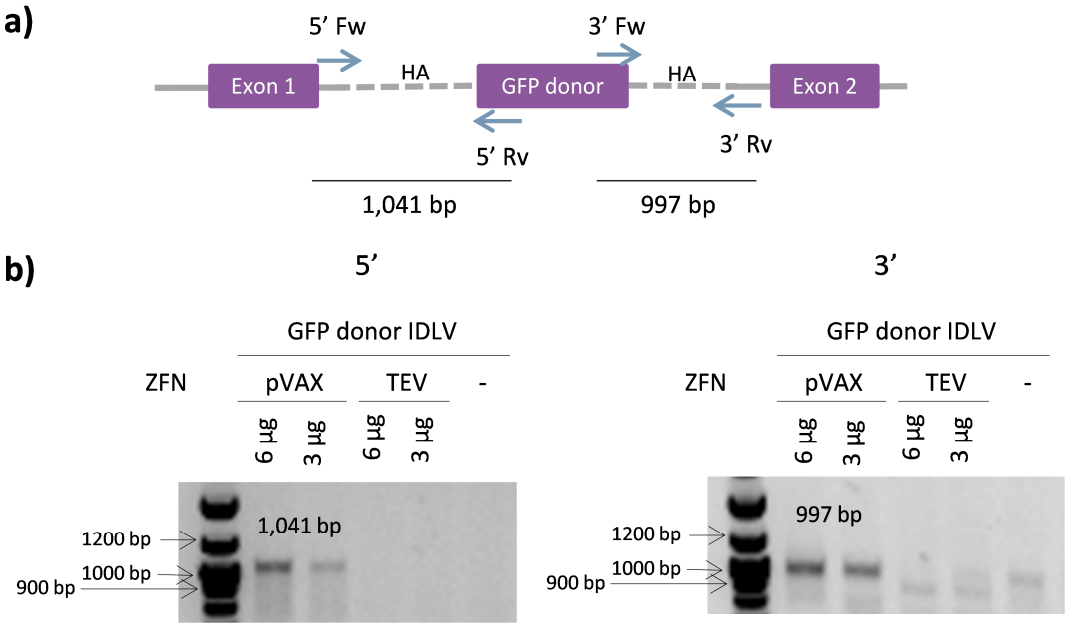


Figure 39: Integration analysis by PCR in gene targeted CD34⁺ cells using different doses of two ZFN constructs*. a) Schematic representation of gene targeted AAVS1 locus with the GFP donor. The arrows represent the primers used to evaluate if integration was site-specific. b) On the left the PCR for the 5' end is shown and on the right the PCR for the 3' end.*

1.2.1.2. Influence of the donor IDLV multiplicity of infection in the gene targeting efficiency in CD34⁺ cells

We also tested if the increment in the MOI of the vector led to an increased efficiency of gene targeting. UCB CD34$^+$ cells were transduced with a GFP donor IDLV at two different MOIs (100 and 200 LVs/cell) and nucleofected with 6 µg of the ZFNs mRNA. The gene targeting efficiency was evaluated at day 14 of culture, measuring the percentage of GFP$^+$ cells by flow cytometry. The results showed that the increase in the donor MOI did not improve the efficiency of gene targeting (Figure 40 a), suggesting that the maximum level of transduction with the IDLV vector was reached. It was also noticeable that the increment in the vector MOI did not increase the random integration in CD34$^+$ cells, as we were not able to detect GFP$^+$ cells when the donor was used alone.

Targeted integration was also evaluated in the bulk population of these samples by PCR. The results showed that gene targeting occurred at both MOIs (Figure 40 b). Taking into account these results we selected an MOI of 100 LVs/cell for the subsequent gene targeting experiments in CD34$^+$ cells.

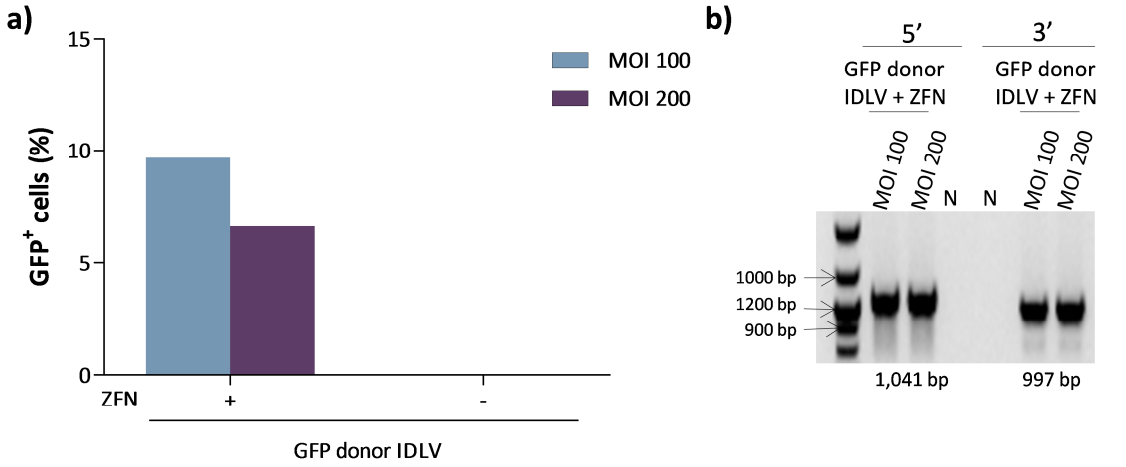


Figure 40: Evaluation of the donor IDLV MOI for gene targeting in CD34⁺ cells. *a) Gene targeting efficiency was evaluated as the percentage of GFP positive cells by flow cytometry in samples targeted with two doses of GFP donor IDLV at day 14 of the protocol. b) On the left the PCR for the 5' end is shown and on the right the PCR for the 3' end. N: negative control.*

1.2.2. Analysis of the gene targeting efficiency with a reporter GFP donor IDLV in CD34⁺ cells

Once all the gene targeting conditions were stablished for human hematopoietic progenitors, eight independent gene targeting experiments were performed in UCB CD34$^+$ cells using a GFP donor IDLV together with ZFNs mRNA to evaluate the feasibility and reproducibility of the technique. Gene targeting efficiency was evaluated by the percentage of GFP$^+$ cells measured by flow cytometry in cells cultured for 7 and 14 days. As shown in Figure 41, GFP$^+$ cells were only detected when the donor and the ZFNs were combined. Mean values of 9.99±4.3% and 10.42±4.92% of gene edited cells were observed in samples cultured for 7 and 14 days respectively, confirming the efficacy of the gene targeting approach.

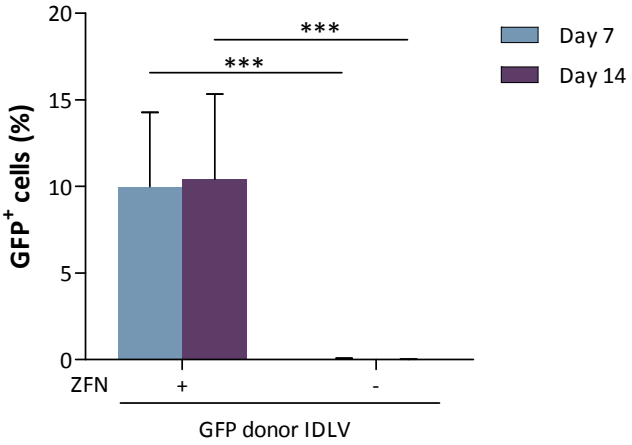


Figure 41: Gene targeting efficiency in HD CD34⁺ cells measured by flow cytometry. *The percentage of GFP$^+$ cells is shown. Bars represent the mean with SD. n=8. The significance of differences between groups is calculated using the T-test, p<0.001 (***).*

Gene targeting efficiency was also measured in different hematopoietic subpopulations analyzed at day 7 and 14 of culture (see section 5.2.1. Materials and methods) (Figure 42 a, b). Although the highest percentage of GFP^+ cells was obtained in the most differentiated population ($CD34^-$) (12.62±4.19% of GFP^+ cells at day 7 of culture); 6.2±4.74% of the primitive progenitors ($CD34^+$, $CD133^+$, $CD90^+$) at day 14 were also GFP^+ cells. Results obtained at day 7 (Figure 42 a) and day 14 (Figure 42 b) of culture showed the same tendency. Notably, no GFP^+ cells were obtained when only the donor IDLV was used, suggesting that random integration did not occur in the different hematopoietic subpopulations.

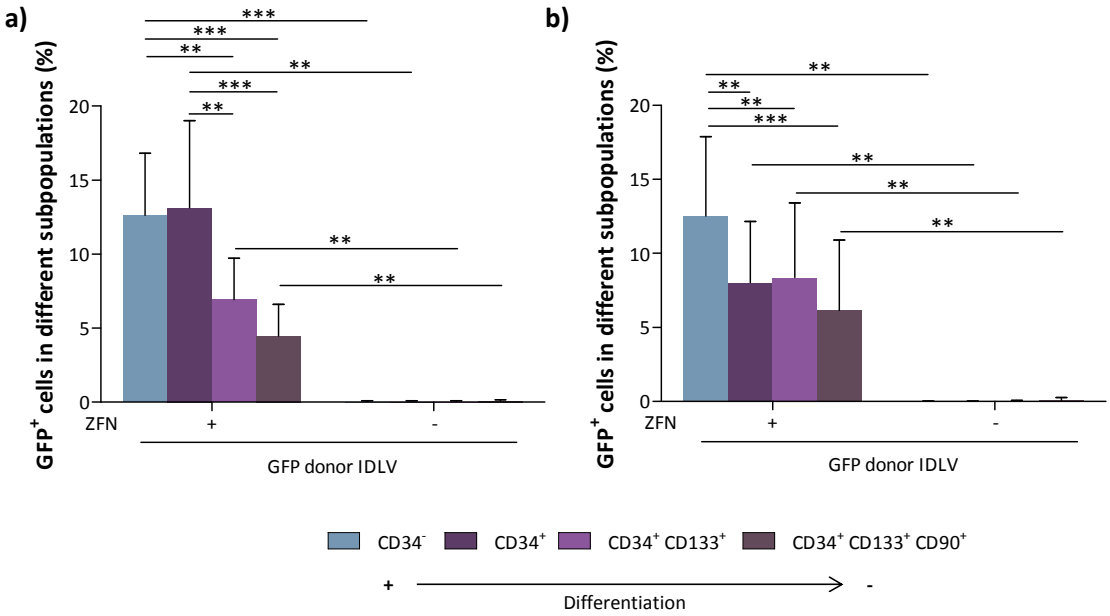


Figure 42: Gene targeting efficiency in different subpopulations of hematopoietic progenitors *(differentiated: $CD34^-$, committed progenitors: $CD34^+$, early progenitors: $CD34^+$ $CD133^+$ and primitive cells: $CD34^+$ $CD133^+$ $CD90^+$). The percentage of GFP^+ cells was measured by flow cytometry after gene targeting. Panel a) represents the analysis at day 7 and panel b) at day 14 of cell culture. Bars represents the mean with SD. n=8. The significance of differences between groups is calculated using ANOVA and T-test and is expressed as p<0.01 (**) and p<0.001 (***).*

Clonogenic assays were conducted with these gene edited cells. Thereafter, GFP positive single colonies were picked when the GFP donor IDLV together with ZFNs were used. When only the GFP donor was used, colonies were picked randomly, since in this case no GFP^+ colonies were detected, and none of these colonies showed the band corresponding to specific integration in the *AAVS1* locus. PCRs both for the 5' and 3' ends were conducted and in 90.48% of the cases the amplification of both ends was confirmed (Figure 43 b and c). In some single colonies the integration was only detected when one of the primers was used, but not the other one, probably due to the different efficiency of the PCR reaction or due to the different mechanisms mediating the repair of each end (HDR or NHEJ)[112,113].

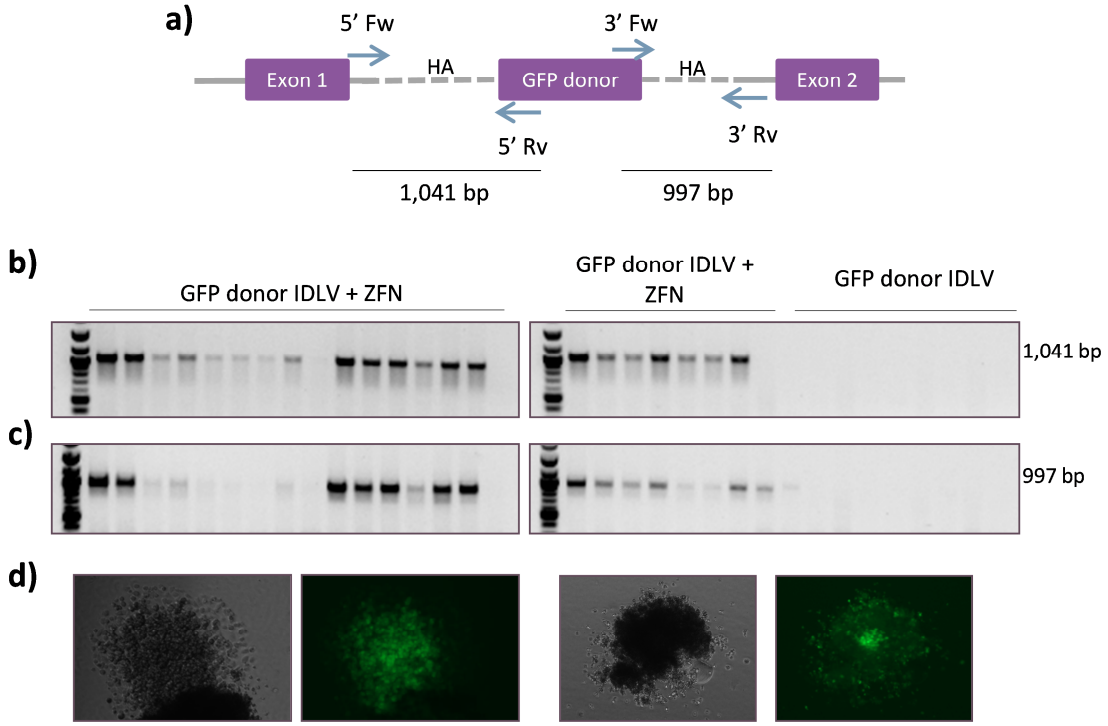


Figure 43: Representative example of the PCR integration analysis in the colonies generated after treatment of CD34$^+$ cells with the GFP donor IDLV together with ZFNs or with the GFP donor IDLV alone. *a) Schematic representation of gene targeted AAVS1 locus with GFP donor. The arrows represent the primers used to evaluate the specific integration of the GFP donor in the AAVS1 site. b) PCR for the 5' end. c) PCR for the 3' end. d) Representative images of two different green colonies obtained in the clonogenic assays when the GFP donor IDLV together with the ZFNs was used. Left colonies showed in bright field and right in fluorescence microscopy.*

The summary of all the experiments performed in HD CD34$^+$ cells is presented in Table 12. The percentage of INDELs quantified in the Surveyor assay varied between 21.36 and 52.85%. The percentage of GFP$^+$ cells measured by flow cytometry at day 14 of cell culture, reached a mean of 10.42%. Since the efficiency of gene targeting analyzed by PCR in GFP$^+$ colonies was 90.48% (confirmed in both 5' and 3' ends), the estimated HDR frequency corresponding to the hematopoietic stem and progenitor cells is 9.43%.

	% INDELs	% GFP$^+$ cells in liquid cultures	% GFP$^+$ in colonies	% HDR in GFP$^+$ colonies
E8	51.65	15.7	24.5	85.7
E9	49.82	9.68	18.06	79.2
E10	42.96	8.23	NA	NA
E11	52.85	18	NA	87.5
E12	48.1	4.87	NA	100
E16	NA	13.5	14.02	NA

	% INDELs	% GFP$^+$ cells in liquid cultures	% GFP$^+$ in colonies	% HDR in GFP$^+$ colonies
E17	21.36	4.4	5.26	100
E19	NA	9	8.82	NA
Mean	**44.46**	**10.42**	**14.13**	**90.48**
SD	**11.83**	**4.91**	**7.58**	**9.22**

Table 12: Summary of the gene targeting experiments in HD CD34$^+$ cells from different donors. *The table shows the percentage of INDELs; proportion of GFP positive cells (liquid culture) measured by flow cytometry at day 14 of cell culture; percentage of GFP$^+$ colonies in the CFC assays in the different experiments; and finally, percentage of HDR in GFP$^+$ colonies, measured by integration analysis of 5' and 3' ends by PCR. E: experiment. NA: not analyzed.*

1.2.2.1. Analysis of the *in vivo* repopulating ability of gene targeted CD34$^+$ cells in immunodeficient mice

As a next step, we evaluated the ability of gene targeted cells to reconstitute the bone marrow (BM) of immunodeficient mice. Cells corresponding to two independent experiments (E10 and 12 shown in Table 12) were used to perform these studies. The efficiency of gene targeting, evaluated as the percentage of GFP$^+$ cells at day 14 of culture, was 8.23% in E10 and 4.87% in E12. The integration analysis performed by PCR in single colonies corresponding to experiment 12 confirmed the specific integration in the *AAVS1* locus for both extremes (Figure 44).

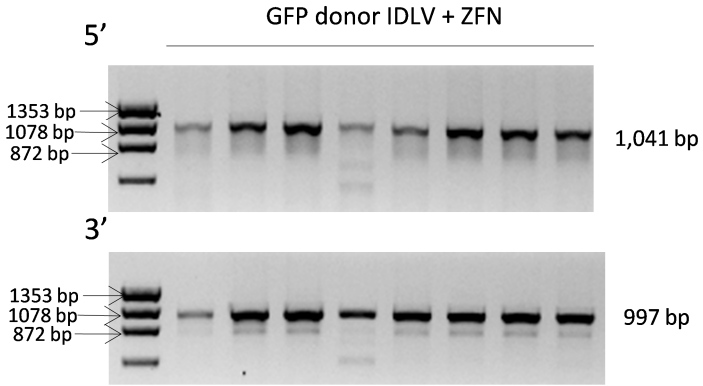


Figure 44: Site-specific integration of the GFP transgene in the AAVS1 site determined in single colonies corresponding to the in vivo experiments. *PCR for the 5' end and PCR for the 3' end in GFP$^+$ colonies assays.*

Aliquots of 4.5x10^5 gene targeted CD34$^+$ cells were transplanted per NSG-SGM3 mouse, pretreated with sublethal irradiation. Flow cytometry analysis of bone marrow (BM) aspirates from transplanted mice revealed both, a high level of human engraftment and also the presence of gene edited cells (represented by hCD45$^+$ cells that express GFP fluorescence (see representative analysis in Figure 45)).

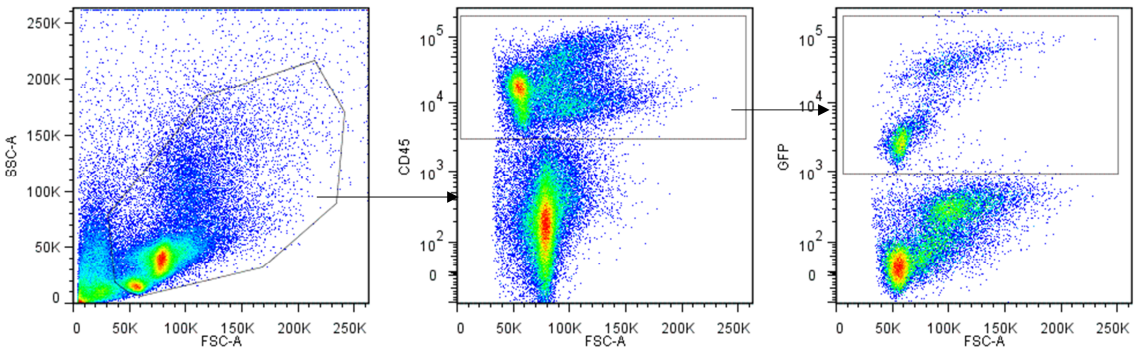


Figure 45: Representative dot plot analysis of the bone marrow cells from a NSG-SGM3 mouse transplanted with gene targeted cells. *From left to right: total BM, hCD45$^+$ cells and finally GFP$^+$ cells are shown.*

Significantly all mice where engrafted with human hematopoietic cells, a median of 17.07% of hCD45$^+$ cells at 90 days post transplantation (dpt Figure 46 a). Moreover the analysis of GFP fluorescence showed that gene edited cells can efficiently engraft in immunodeficient mice, even at 90 days post-transplantation. At this time a median efficiency of 11.93% human CD45$^+$ cells were GFP$^+$, fluctuating between 0.2 and 28.3% of GFP positive cells depending on the mouse (Figure 46 b). Notably, in most cases the percentage of gene edited cells was maintained or increased at 90 dpt, suggesting that gene targeting has occurred at the level of hematopoietic repopulating cells.

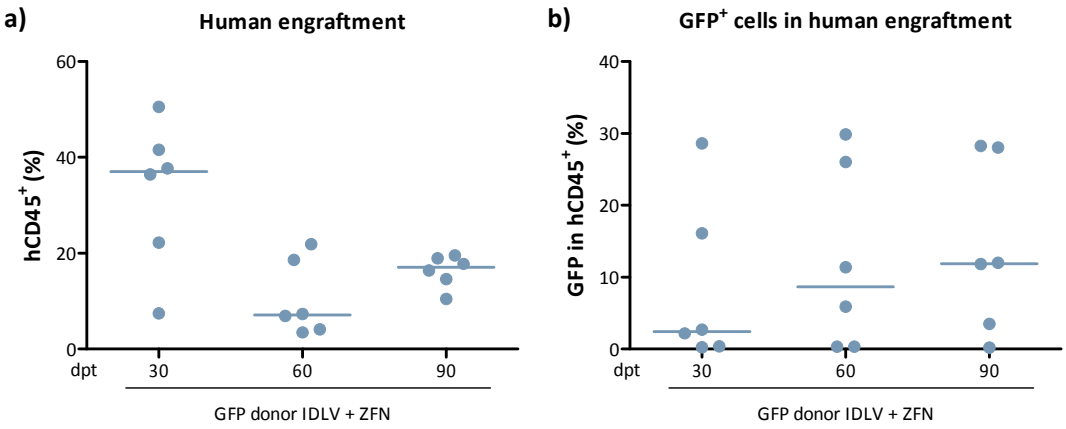


Figure 46: Analysis of the engraftment of gene targeted cells in NSG-SGM3 mice. *a) Human engraftment (% hCD45$^+$ cells) 30, 60 and 90 days post-transplantation (dpt). b) GFP$^+$ population in human hematopoietic cells at 30, 60 and 90 dpt. Horizontal lines represent the median. n=6.*

Multilineage analysis performed in the BM and spleen of these animals at 90 dpt showed that there was human engraftment in all hematopoietic subpopulations in different hematopoietic organs (34.65 - 64.35% human engraftment in spleen) (Figure 47 a and c). We were also able to detect GFP positive cells in all hematopoietic lineages in BM and spleen of recipients transplanted with gene edited cells, reaching values of 8.95% GFP positive cells in spleen (median of 4.47%). The multilineage engraftment of GFP$^+$ cells confirmed that gene editing occurred in HSPCs and did not affect the differentiation capacity of these hematopoietic precursors (Figure 47 b and d).

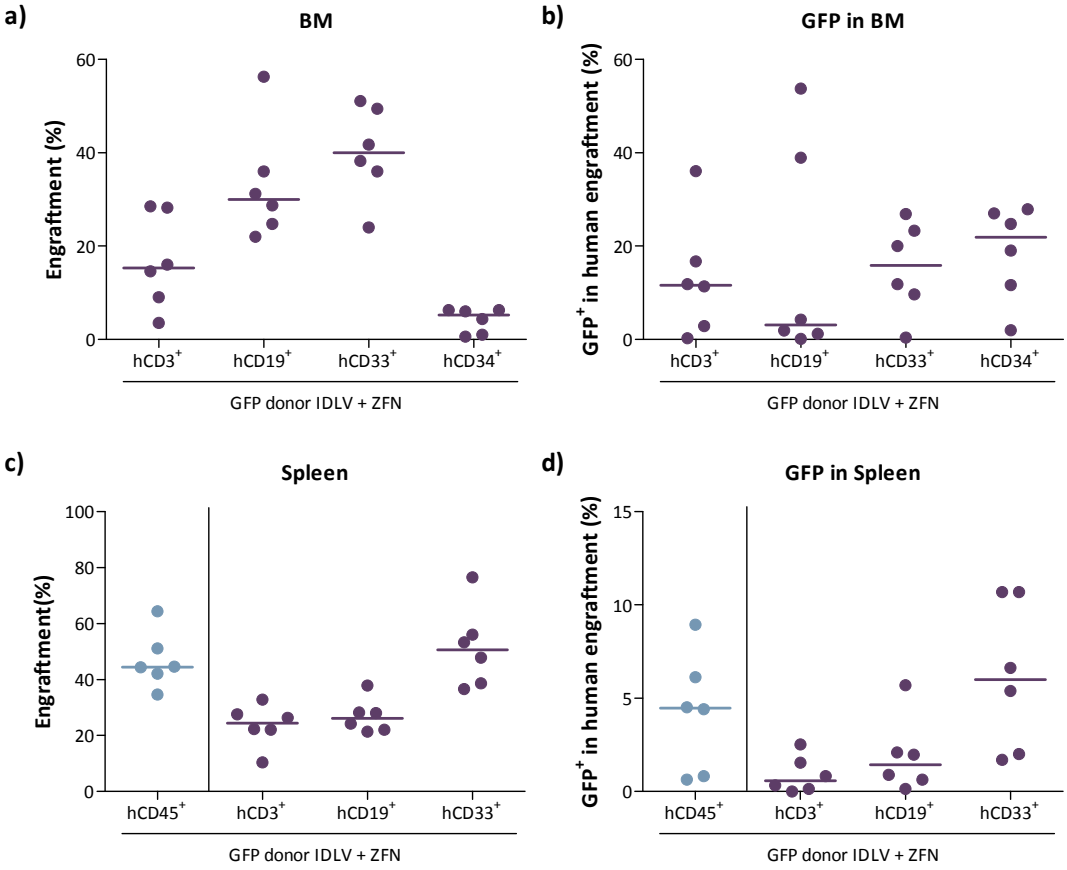


Figure 47: Multilineage engraftment of gene edited cells in different hematopoietic organs. *Multilineage reconstitution was evaluated by flow cytometry using antibodies anti- hCD3$^+$ for T cells, hCD19$^+$ for B cells, hCD33$^+$ for myeloid cells and hCD34$^+$ for progenitor cells. a) Multilineage engraftment in the BM. b) Percentage of GFP$^+$ edited cells in the different subpopulations in the BM. c) Multilineage engraftment in spleen. The first column represents the total human engraftment in this organ. d) Percentage of GFP$^+$ cells in the different subpopulations in the spleen. The first column represents the total content of GFP$^+$ cells in hCD45$^+$ in this organ. Horizontal lines represent the median. n=6.*

1.2.2.1.1. Analysis of the long-term repopulating capacity of gene targeted CD34$^+$ cells

To evaluate if gene targeting was occurring at the level of hematopoietic stem cells, bone marrow from primary recipients was harvested. Human CD45$^+$ cells purified by immunoselection were seeded in methylcellulose to evaluate their clonogenic capacity.

Analysis of the specific integration by PCR was conducted in GFP positive colonies detected by fluorescence microscopy in the clonogenic assays conducted with cells from primary mice (Figure 48 a). The integration analysis of these colonies by PCR showed the site-specific integration of the *GFP* transgene (Figure 48 b), confirming that gene edited cells were stably maintained in primary recipients.

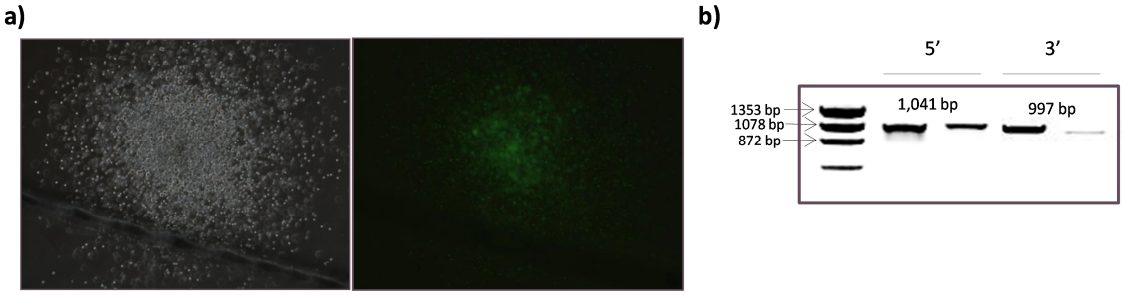


Figure 48: Analysis of the site-specific integration in CFCs obtained from primary mice transplanted with gene edited cells. *a) Representative image of a GFP$^+$ colony. Left panel shows a colony under bright-field microscopy and left panel the same colony under fluorescence microscopy. b) PCR analysis of two different GFP$^+$ colonies obtained from primary recipients to detect targeted integration in the AAVS1 locus.*

1.2.3. Analysis of the gene targeting efficiency using the GFP-FANCA donor IDLV in CD34$^+$ cells from healthy donors

To test the efficiency of a therapeutic vector to specifically integrate in the *AAVS1* locus of HSPCs and engraft in immunodeficient mice, we initially focused our efforts to verify integration of the GFP-FANCA donor in HD CD34$^+$ cells. As we previously observed in HD LCLs we could not detect GFP expression by flow cytometry (Figure 49 a). To estimate the gene targeting efficiency with this vector we picked randomly single colonies generated after the transfer of the GFP-FANCA donor IDLV and ZFNs in CD34$^+$ cells. First we did a Q-PCR to detect the number of *GFP* copies in these single colonies. This analysis showed that 19.5% of these cells presented more than 0.5 copies of *GFP*. To verify that in these colonies the integration of the vector was site-specific we performed a PCR to detect the integration in a pull of colonies with more than 0.5 copies of the GFP vector per cell. Our results showed the specific integration of the donor in the *AAVS1* locus by PCR analysis in the 5' end (Figure 49 b), confirming that targeting HD CD34$^+$ cells with this therapeutic donor is feasible.

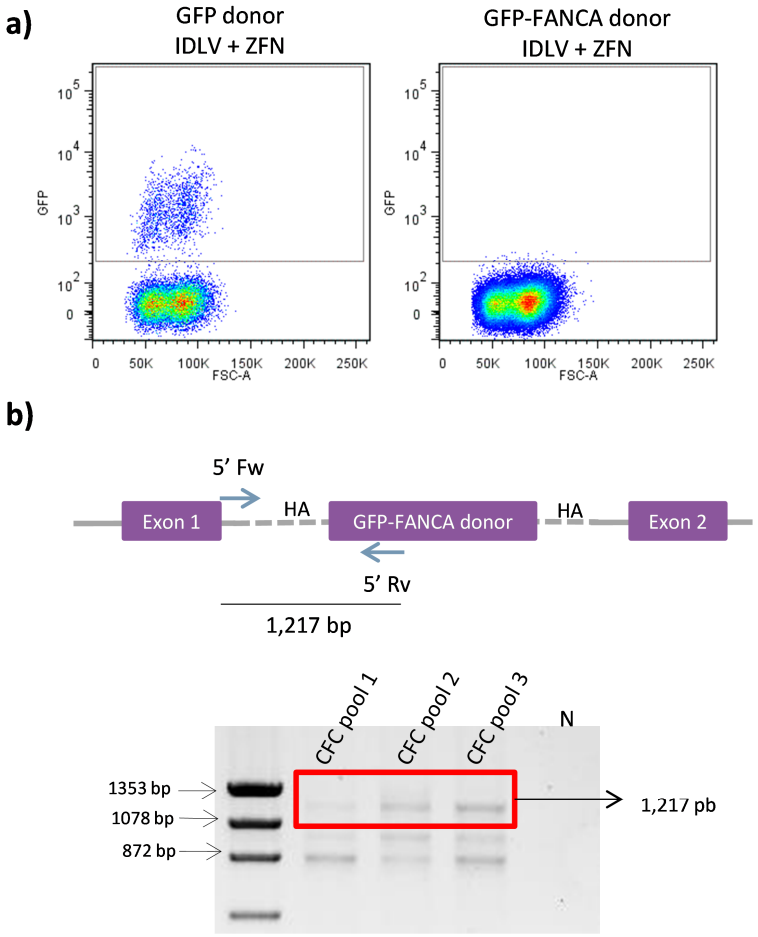


Figure 49: Gene targeting of the GFP-FANCA donor in the AAVS1 locus of HD CD34$^+$ cells. *a) Representative flow cytometry dot plot of CD34$^+$ cells targeted with the GFP donor or GFP-FANCA donor at day 14 of culture. b) Top: Schematic representation of gene targeted AAVS1 locus with GFP donor. The arrows represent the primers used to evaluate the specific integration of the donor in the AAVS1 site. Bottom: Integration analysis by PCR of different pools of individual colonies that presented more than 0.5 copies of GFP/cell determined by Q-PCR analysis. N: negative control.*

Since gene targeting with GFP-FANCA donor was confirmed in CD34$^+$ cells, cells transduced with this donor were transplanted in immunodeficient NSG-SGM3 mice. Once again, our results showed the engraftment of human hematopoietic cells in primary recipients (Figure 50). Moreover, the analysis of GFP VCN in these samples by Q-PCR confirmed the engraftment of genetically modified cells.

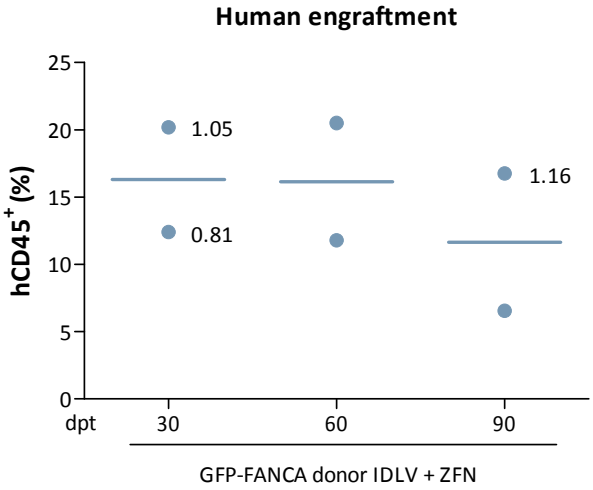


Figure 50: Analysis of human engraftment of gene targeted cells using donor GFP-FANCA in primary NSG-SGM3 recipients. *Human engraftment (% hCD45$^+$ cells) 30, 60 and 90 dpt in primary recipients. Numbers represent VCN analyzed by Q-PCR in each sample. Horizontal lines represent the median. n=2.*

The multilineage analysis of primary recipients showed the repopulation of the lympho-hematopoietic lineages in BM and spleen (Figure 51). Moreover, the GFP VCN analysis performed in the spleen of these mice also confirmed the engraftment of gene edited cells.

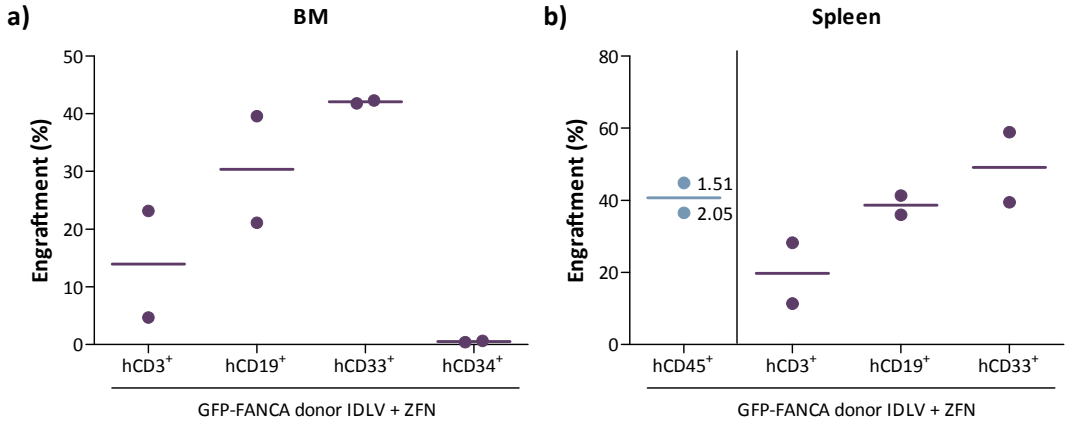


Figure 51: Multilineage engraftment of primary recipients transplanted with human CD34$^+$ cells after gene editing. *Recipients transplanted with CD34$^+$ cells targeted with the GFP-FANCA donor IDLV. The engraftment was evaluated with antibodies anti-hCD3$^+$ for T cells, hCD19$^+$ for B cells, hCD33$^+$ for myeloid cells and hCD34$^+$ for progenitor cells. a) Multilineage engraftment in BM. b) Multilineage engraftment in spleen. Numbers represents VCN analyzed by Q-PCR in each sample. Horizontal lines represent the median. n=2.*

1.3. Analysis of the efficiency of gold nanoparticles to improve lentiviral transduction and gene targeting in CD34$^+$ cells

One essential aspect to improve the efficiency of HDR is to improve the delivery of the donor constructs into the target cell. We investigated whether the transduction efficiency in CD34$^+$

cells from UCB could be improved using gold nanoparticles (AuNPs) conjugated with NH_4^+ residues. These AuNPs have a positive charge, and therefore could be bound to negatively charged LVs. Our hypothesis was that the inhibition of the negative charge of LVs could improve their efficiency to transduce the target cell population, in particular the $CD34^+$ cells, improving the delivery of the donor construct.

First we tested the toxicity of these particles in $CD34^+$ cells, and also if these cells were able to incorporate them efficiently. We analyzed the number of cells one and two days after treatment. As shown in Figure 52 a, although one day after treatment with AuNPs a decrease in the number of cells was obtained, this effect was not significant. Moreover, at two days after treatment no differences in $CD34^+$ cell number, cultured with or without AuNPs, were observed. Strikingly, we demonstrated by fluorescence microcopy that $CD34^+$ cells incorporated the AuNPs very efficiently (Figure 52 b). On average, more than 90% of the $CD34^+$ cells contained AuNPs.

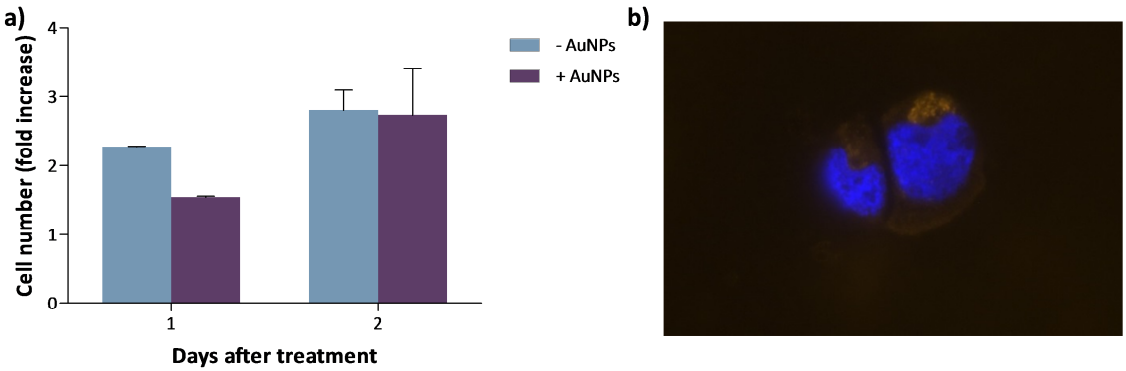


Figure 52: Analysis of the proliferation and incorporation of gold nanoparticles into the $CD34^+$ cells. *a) Fold increase in the number of cells is represented. It is determined by the cell number observed one day over the number of cells counted the day before. Bars represent mean with SD. n=3. b) Image of $CD34^+$ cells in which the nanoparticles have been incorporated. Nucleus in blue (DAPI) and AuNPs in orange (TAMRA).*

In order to test if the transduction efficiency of LVs could be improved by the binding to AuNPs, $CD34^+$ cells were transduced with LVs pre-incubated or not with AuNPs. Different MOIs of the GFP donor IDLVs or GFP donor ICLVs were used. The results showed that the pre-incubation of either IDLVs or ICLVs with AuNPs markedly increased the transduction efficiency of these vectors. On average a two fold increase was observed (Figure 53).

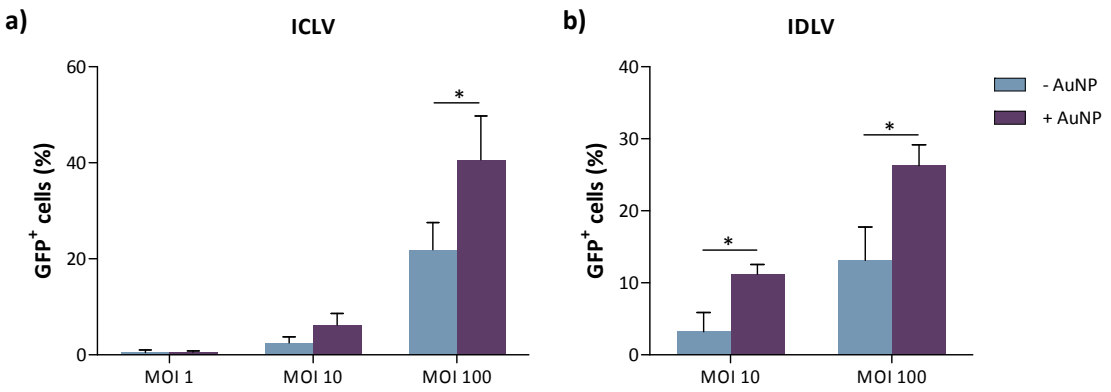


Figure 53: Transduction efficiency in UCB CD34$^+$ cells using a GFP donor ICLV or a GFP donor IDLV in the presence of AuNPs. *a) The percentage of GFP positive cells transduced with a GFP donor ICLV was measured 12 days after transduction. b) Percentage of GFP positive cells transduced with a GFP donor IDLV measured at 2 days post-transduction. Bars represent mean with SD. n=4. The significance of differences between groups is calculated using Mann Whitney test and is expressed as p<0.05 (*).*

We also tested if the transduction improvement associated to the use of AuNPs in *in vitro* cultured CD34$^+$ cells was correlated with similar improvements in NSG-repopulating cells. In these experiments, CD34$^+$ cells were transduced with two doses of donor GFP ICLV (MOI 10 or 100 LVs/cell) in the presence or the absence of AuNPs. Thereafter, cells were transplanted into sublethally irradiated NSG mice. We decided to use two different MOIs as MOI 100 LVs/cell could result in very high transduction of repopulating cells that may limit the detection of statistical differences. The human engraftment of these animals was evaluated monthly by BM aspiration until 90 dpt. Our results showed that the percentage of human engraftment did not change because of the use of AuNP-LVs at any time post-transplantation, indicating that AuNPs did not affect the capacity of HSPCs to repopulate the BM of immunodeficient mice, confirming the absence of toxicity of these AuNPs (Figure 54 a). When the percentage of GFP$^+$ cells engrafting the animals was analyzed, we observed a slight increase in the percentage of transduced cells when AuNPs were used, although differences were only statistically significant in the short-term repopulating cells when MOI 100 was used (three-fold increase in GFP$^+$ hCD45$^+$ cells when AuNPs were used) (Figure 54 b). When the percentage of GFP$^+$ cells was analyzed in the CD34$^+$ subset we observed the same tendency, cells transduced in the presence of AuNPs presented an increment of this percentage, especially long term post-transplantation (Figure 54 c). As expected, the higher MOIs led to a higher percentage of GFP$^+$ cells in transplanted mice. This result indicated that although the efficiency of transduction was markedly improved in the liquid culture, this increase is mainly detected in the short-term *in vivo*, and only a mild increase is observed long term post-transplantation for both MOIs used (Figure 54 b and c).

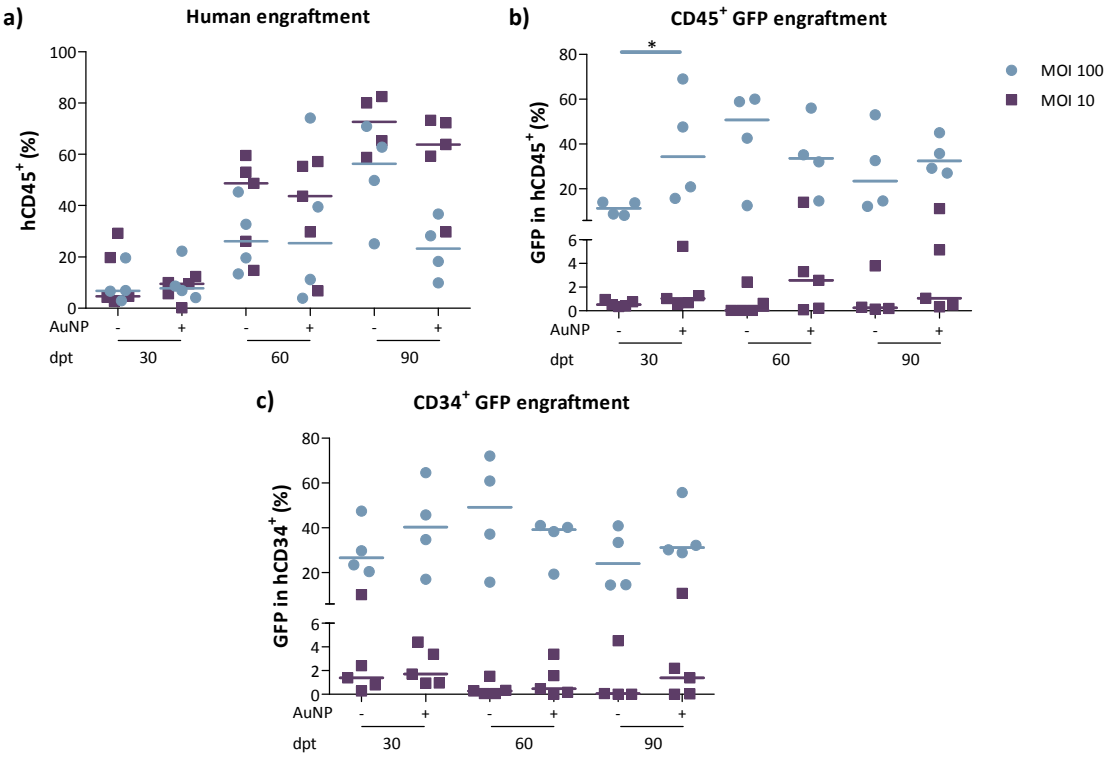


Figure 54: Engraftment capacity of CD34$^+$ cells transduced in the presence of AuNPs in NSG mice. *Cells were transduced with two different MOIs. Circles represent cells transduced with MOI 100 LVs/cell, while squares represent cells transduced with MOI 10 LVs/cell. a) Human engraftment (% hCD45$^+$ cells) 30, 60 and 90 dpt. b) GFP$^+$ population in the human engraftment at 30, 60 and 90 dpt. c) GFP$^+$ cells inside CD34$^+$ subpopulation at 30, 60 and 90 dpt. Horizontal lines represent the median. n=4 in MOI 100 and n=5 in MOI 10. The significance of differences between groups is calculated using Mann Whitney test and is expressed as p<0.05 (*).*

The multilineage analysis of the hematopoietic organs of these mice at 90 dpt showed that the distribution of human hematopoietic lineages in infused recipients was very similar between the different conditions. The lymphoid, myeloid and progenitor compartments were represented at the same levels in the different analyzed organs in the conditions treated or not with AuNPs (Figure 55 a, c and e). Moreover, the expression of GFP in all these different compartments was similar in animals transplanted with cells transduced with or without AuNPs; although there was a slight increment of GFP$^+$ cells in all hematopoietic compartments of mice analyzed when AuNPs were used, especially in the thymus in which the human reconstitution was around 100% (Figure 55 f). As expected we observed that animals transplanted with cells transduced with higher MOIs show higher levels of GFP$^+$ cells (Figure 55 b, d and f).

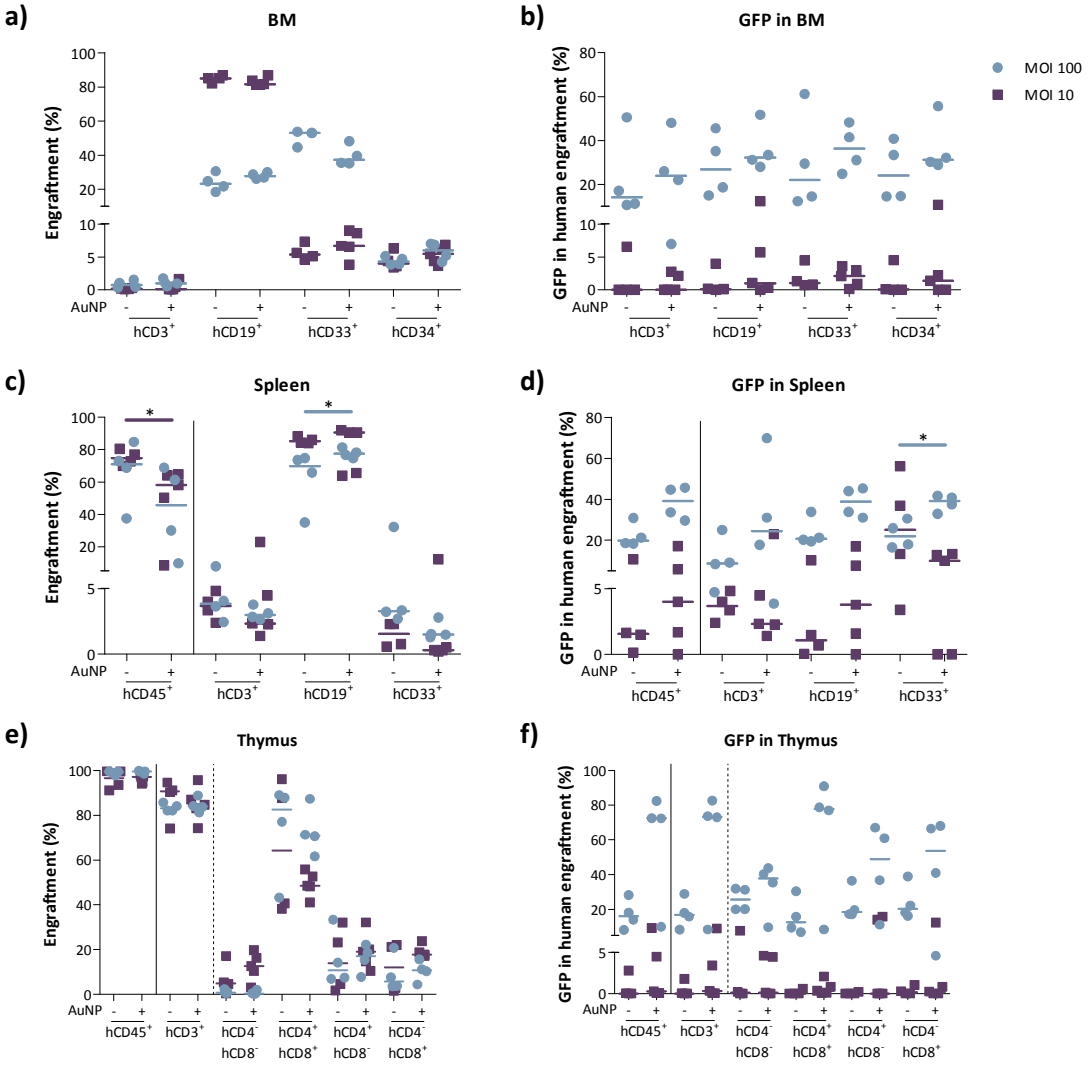


Figure 55: Multilineage engraftment of transduced cells with or without AuNPs transplanted in NSG mice. *Multilineage reconstitution was evaluated using antibodies against hCD3$^+$ for T cells, hCD19$^+$ for B cells, hCD33$^+$ for myeloid cells, hCD34$^+$ for progenitor cells and with hCD4$^+$ and hCD8$^+$ for different subpopulations of T cells in the thymus. Circles represent cells transduced with MOI 100 LVs/cell, while squares represent cells transduced with MOI 10 LVs/cell. a) Multilineage engraftment in the BM. b) Percentage of GFP edited cells in the different subpopulations in the BM. c) Multilineage engraftment in the spleen. d) Percentage of GFP$^+$ cells in the different subpopulations inside the spleen. e) Analysis of T cell populations in the thymus. f) Percentage of GFP$^+$ cells in the different subpopulations of T cells in the thymus. Horizontal lines represent the median. n=4 in MOI 100 and n=5 in MOI 10. The significance of differences between groups was calculated using Mann Whitney test and is expressed as p<0.05 (*).*

Finally, secondary transplants were performed using the BM of the primary recipients. The analysis of the secondary recipients showed that there is a moderate increment in the engraftment of GFP$^+$ cells when nanoparticles are used at both MOIs, although these differences were not statistically significant (Figure 56).

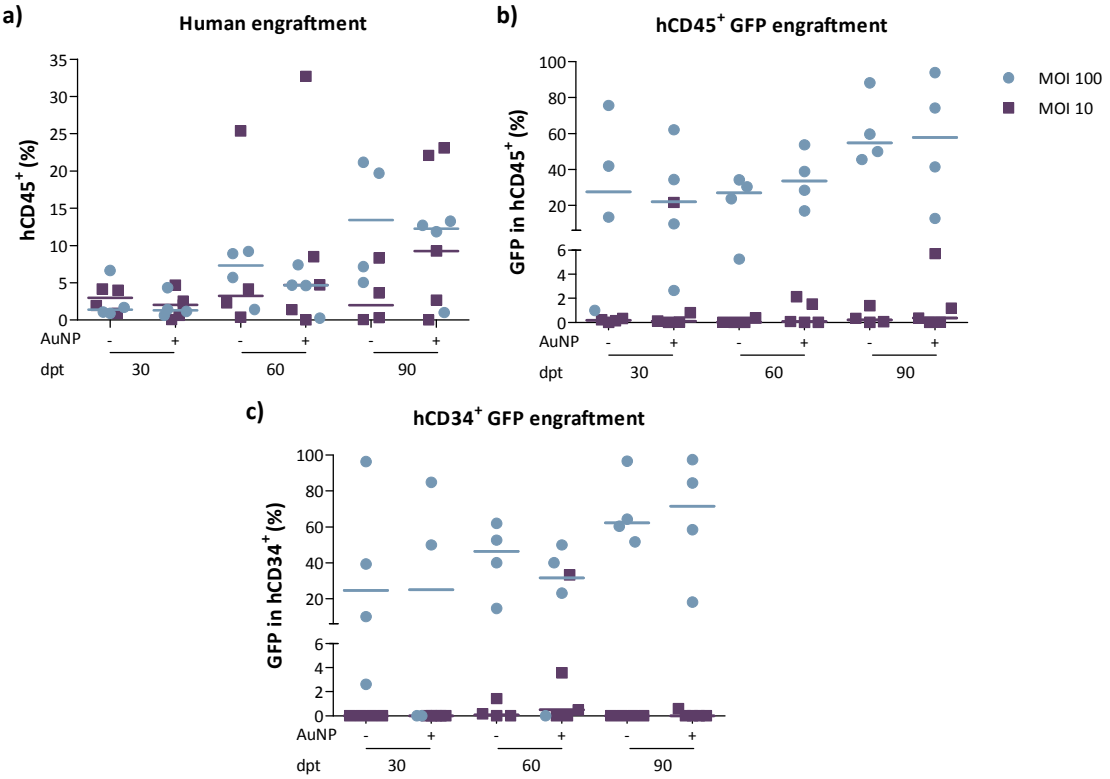


Figure 56: Analysis of the engraftment of CD34$^+$ cells transduced in the absence/presence of AuNPs in secondary recipients. *Cells were transduced with two different MOIs. Circles represent cells transduced with MOI 100 LVs/cell, while squares represent cells transduced with MOI 10 LVs/cell. a) Human engraftment (% CD45$^+$ cells) 30, 60 and 90 dpt. b) GFP$^+$ population inside of human engraftment at 30, 60 and 90 dpt. c) GFP$^+$ cells inside CD34$^+$ subpopulation at 30, 60 and 90 dpt. Horizontal lines represent the median. n=4 in MOI 100 and n=5 in MOI 10.*

The multilineage analysis of all hematopoietic organs of these mice at 90 dpt showed that there were no differences in the lineage reconstitution, since the lymphoid, myeloid and progenitor compartments were represented at the same levels in the different organs (Figure 57 a, c and e). Moreover, the expression of GFP in all these different compartments was similar in the animals transplanted with cells transduced with or without AuNPs. As expected, we also observed that animals with higher percentages of GFP positive cells were the ones transplanted with cells transduced with higher MOIs (Figure 57 b, d and f).

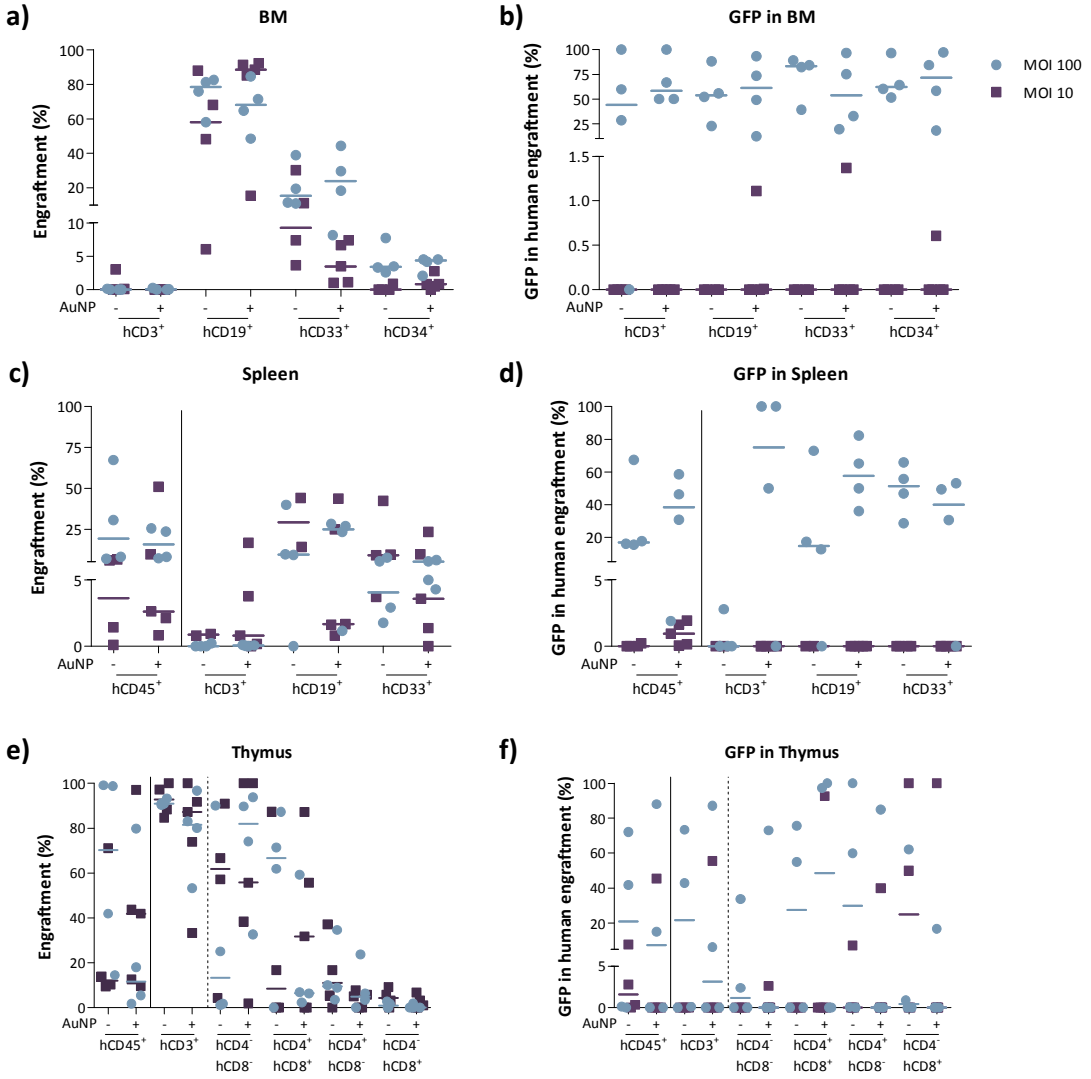


Figure 57: Multilineage engraftment of transduced cells with or without AuNPs transplanted in NSG mice. *Multilineage reconstitution was evaluated using antibodies against hCD3$^+$ for T cells, hCD19$^+$ for B cells, hCD33$^+$ for myeloid cells, hCD34$^+$ for progenitor cells and wit hCD4$^+$ and hCD8$^+$ for different subpopulations of T cells in the thymus. Circles represent cells transduced with MOI 100 LVs/cell, while squares represent cells transduced with MOI 10 LVs/cell. a) Multilineage engraftment in the BM. b) Percentage of GFP edited cells in the different subpopulations in the BM. c) Multilineage engraftment in the spleen. d) Percentage of GFP$^+$ cells in the different subpopulations inside the spleen. e) Analysis of T cell populations in the thymus. f) Percentage of GFP$^+$ cells in the different subpopulations of T cells in the thymus. Horizontal lines represent the median. n=4 in MOI 100 and n=5 in MOI 10.*

Finally the vector copy number was analysed in the BM from primary and secondary recipients at 90 dpt (Figure 58). Although the treatment with AuNPs did not increase the number of copies found in primary recipients (Figure 58 a); in the secondary recipients the treatment with gold nanoparticles increased the number of copies in the group transduced with MOI 100 LVs/ cells (from 1.47 to 3.26 copies/cell). Moreover, while we did not detect copies of the transgene in animals corresponding to the MOI 10 group, we could detect copies of the LV when cells were treated with AuNPs (0.089 copies/cell) (Figure 58 b).

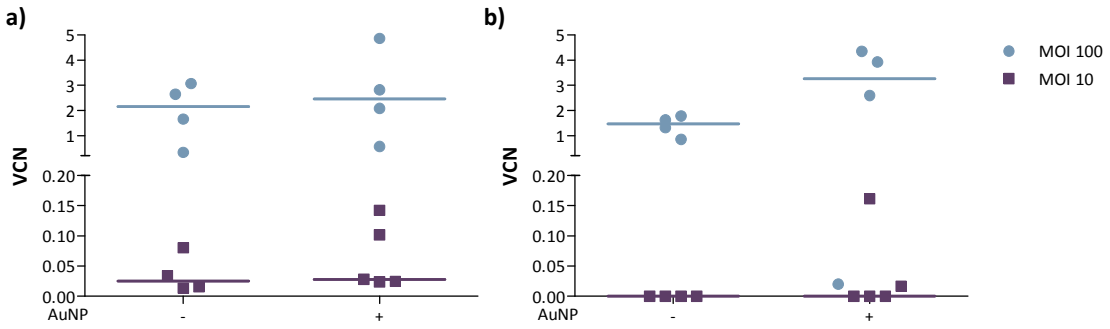


Figure 58: Vector copy number analysis in human cells obtained from the BM of primary and secondary recipients at 90 dpt. *a) VCN analysis in primary recipients. b) VCN analysis in secondary recipients. Horizontal lines represent the median. n=4 in MOI 100 and n=5 in MOI 10.*

Finally, and considering the significant increase obtained in the percentage of transduction when IDLVs were used (Figure 53 b), we tested the efficiency of gene targeting when AuNPs were combined with IDLVs in $CD34^+$ using the GFP donor IDLV. Our results showed that the combination of AuNPs in cells after transduction with IDLVs and the nucleofection led to a high toxicity in $CD34^+$ cells, preventing the evaluation of the effect of AuNPs in gene targeting experiments (data not shown).

2. Gene targeting with a therapeutic donor IDLV in FA-A hematopoietic cells

2.1. Evaluation of gene targeting efficiency in FA-A LCLs

Although our previous studies demonstrated that gene targeting of h*FANCA* can be conducted in FA-A fibroblasts[75], the efficiency of this approach in FA-A hematopoietic cells, such as LCLs, in comparison to HD LCLs has not been tested. Since the FA/BRCA pathway is involved in HR[228], we wanted to examine if FA-A hematopoietic cells show a reduction in the gene targeting efficiency as compared to HD LCLs. To roll out any potential effect mediated by the transient expression of FANCA, a GFP donor IDLV was used in these experiments. Cell lines from four different FA-A patients, both uncorrected and corrected with a therapeutic lentiviral vector[290], were used (Figure 59 b). Additionally, four HD LCLs were also included (Figure 59 a). Our results showed that gene targeting was feasible in FA-A hematopoietic cells and the efficiency of gene targeting was dependent on the specific cell line. Re-expression of h*FANCA* gene in FA-A LCLs from patients mildly increased the gene targeting efficiency in three out of four analyzed cell lines. However, there were not significant differences between HD and FA-A cells corrected or not. We also corroborated that the integration was site-specific by PCR analysis of both 5' and 3' ends of *AAVS1* locus in GFP^+ cells obtained from these experiments by cell sorting (Figure 59 c).

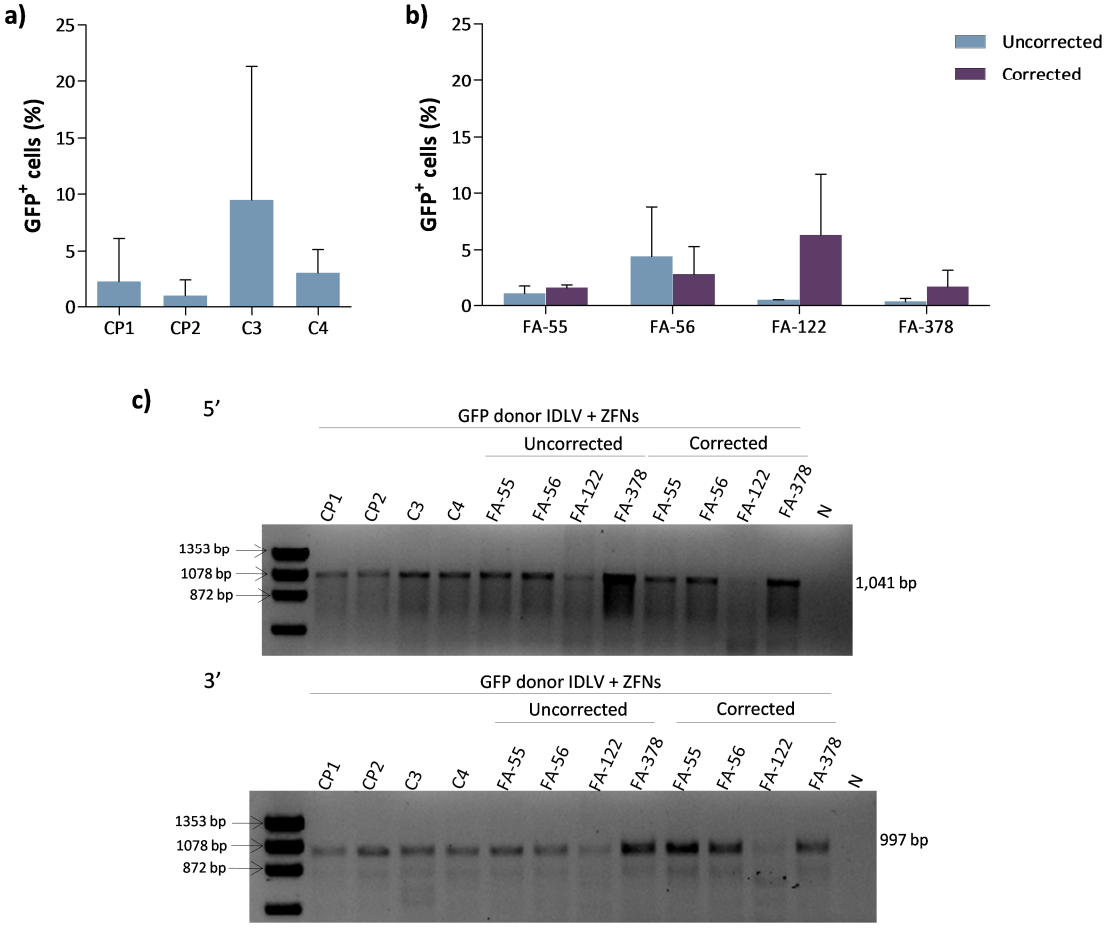


Figure 59: Analysis of the gene targeting efficiency in different LCLs. *The percentage of GFP$^+$ cells measured by flow cytometry at day 12 of culture from gene targeting experiments using GFP donor IDLV is represented. a) Percentage of GFP$^+$ cells from four different HD LCLs after gene targeting. b) Percentage of GFP$^+$ cells obtained in gene targeting experiments from four different FANCA deficient LCLs and their gene therapy corrected counterparts. Bars represent mean with SD. n=3. c) Representative PCR of the 5' end (top) and 3' end (bottom) of the sorted bulk population of gene targeted HD, uncorrected and corrected FA-A LCLs using GFP donor IDLV and ZFNs. N: negative control.*

2.2. Gene targeting of *FANCA* in the *AAVS1* locus of FA-A lymphoblastic cell lines

2.2.1. Design of a new therapeutic vector to correct FA-A hematopoietic cells

Due to the difficulties to detect the GFP expression conferred by the integration of the GFP-FANCA donor in the *AAVS1* locus, we designed a new IDLV carrying the h*FANCA* gene and the puromycin (Puro) resistant gene, both of them under the control of the PGK promoter. To test the functionality of this vector we initially produced it as an integrase-competent lentiviral vector (ICLV) and analyzed its efficiency to revert the phenotype of FA-A LCLs. In all these experiments the GFP donor vector, also produced as ICLV, was used as a control.

A lymphoblastic cell line from a FA-A patient (FA-55) was transduced with a MOI 10 LVs/cell in two rounds of 12 hours of transduction with both vectors. The percentage of GFP positive cells

was evaluated by flow cytometry when the GFP donor ICLV was used, obtaining a 39.5% of GFP^+ cells 15 days after transduction. To confirm the functionality of the therapeutic FANCA-Puro vector four approaches were conducted in these cells: reversion of MMC sensitivity, western-blot analysis of FANCA and generation of FANCD2 foci formation after DNA damage. The resistance to puromycin was also tested.

Fifteen days after transduction cells were plated with different concentrations of MMC or puromycin. The evaluation of MMC and puromycin sensitivity was performed by analyzing the percentage of DAPI negative cells at 15 days post-treatment by flow cytometry. The results showed that the transduction of the FA-A LCL with our FANCA-Puro therapeutic vector reverts the sensitivity of these cells to MMC (Figure 60 a), and also confers resistance to puromycin (Figure 60 b), confirming that both transgenes were functional in the vector.

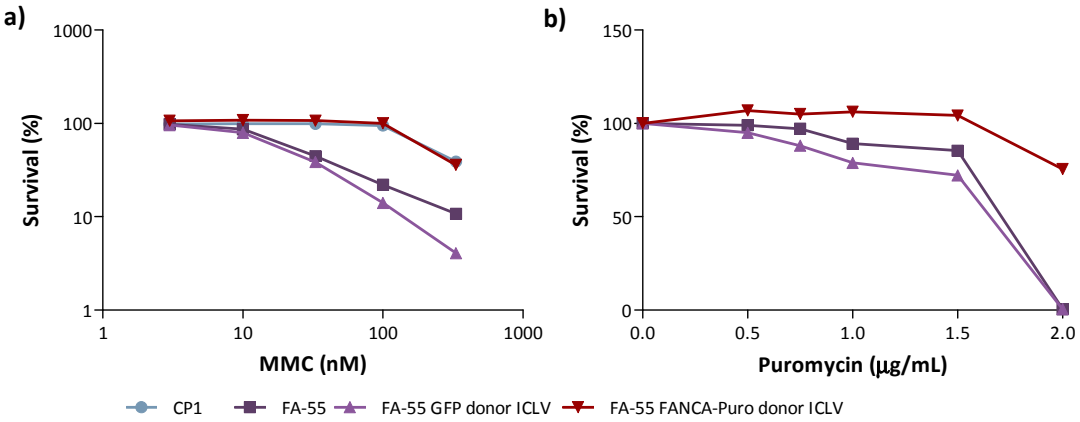


Figure 60: Survival curve of FA-55 LCL after transduction with the GFP or FANCA-Puro ICLVs. *a) MMC survival curve of transduced cells. b) Puromycin survival curve of transduced cells. HD cell line CP1 and FA55 were used as controls.*

To confirm the expression of FANCA in these cells, western blot analysis of FANCA were conducted. As shown in Figure 61 FANCA expression was only detected in samples transduced with the FANCA-Puro vector, either unselected or after selection with MMC or puromycin, demonstrating the efficiency of the vector to promote the expression of the therapeutic transgene.

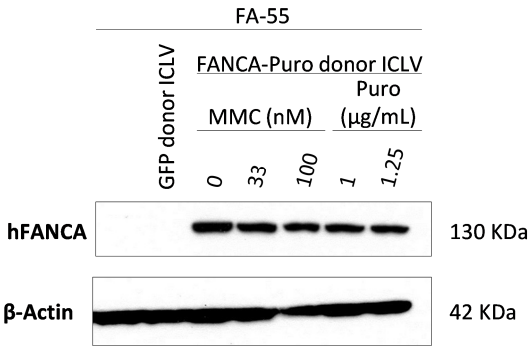


Figure 61: Western blot analysis of FANCA in FA-55 edited LCLs. *β-actin was used as a loading control.*

Since FANCA is necessary for FANCD2 monoubiquitination and mobilization to the DNA repair foci, we investigated the generation of FANCD2 foci after DNA damage in FANCA-Puro

transduced cells. As shown in Figure 62 only in FA-A cells that had been transduced with the FANCA-Puro vector the presence of FANCD2 foci was noted.

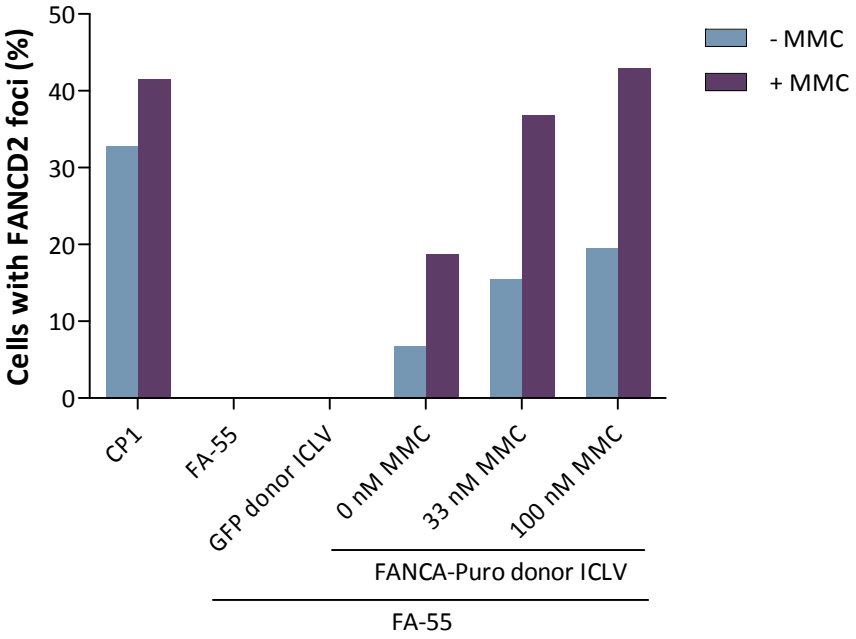


Figure 62: Percentage of cells with FANCD2 foci in a FA-A LCL after gene targeting using a FANCA-Puro donor ICLV or GFP donor ICLV. *Cells with more than 10 foci were counted.*

Taken together, all these studies showed the capacity of the FANCA-Puro vector to restore the phenotype of a FA-A LCL. In subsequent experiments we thus investigated the efficiency of this construct to specifically conduct gene targeting in the *AAVS1* locus of FA-A hematopoietic cells.

2.2.2. Gene targeting efficiency of the donor FANCA-Puro delivered as plasmid DNA in LCLs

In these experiments a FA-A LCL (FA-55) was nucleofected either with the donor FANCA-Puro construct or a control GFP donor, both delivered as plasmid DNA. A HD LCL (C4) was used as a control. Flow cytometry analyses performed at day 12 of culture in samples targeted with GFP donor showed that there were GFP^+ cells in samples only nucleofected with the donor as DNA in both HD and FA-A LCLs. Notably, the percentage of positive cells was markedly higher in FA-A cells nucleofected only with the donor in comparison with the HD, pointing out the possibility that random integration could be favored in FA-A cells due to its intrinsic genetic instability (Figure 63).

It is important to remark that the percentages of INDELs obtained in both LCLs and using both donor vectors (GFP and FANCA-Puro) were very similar: around 50% (49.19±3.18%).

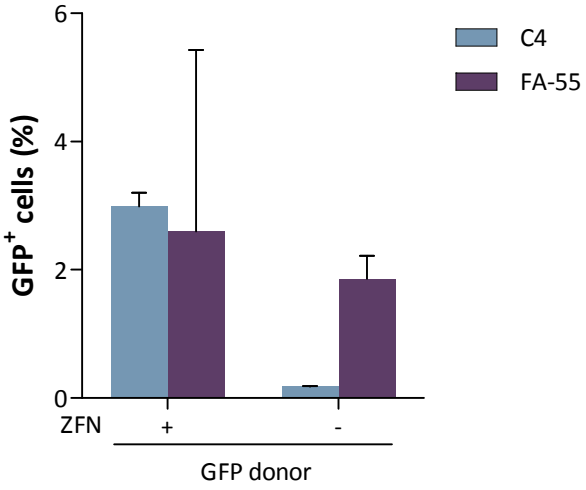


Figure 63: Gene targeting in FA-55 LCLs using GFP donor as DNA. *Percentage of GFP positive cells measured by flow cytometry in FA-55 LCL cell line and C4 HD at day 12 of culture. Bars represent mean with SD. n=2.*

The integration analysis in the 5' end of the *AAVS1* locus of FA-55 LCLs targeted using GFP or FANCA-Puro donor as plasmid DNA confirmed the specific integration of the donor FANCA-Puro, but not the donor GFP (Figure 64 b). The absence of the PCR amplicon in cells targeted with the GFP donor could be due to the low percentage of GFP^+ cells obtained in these experiments (2.59±2.84%), which were very close to the detection limit of this PCR (2.25% of GFP^+ cells; see Figure 35). However, although the size of the FANCA-Puro donor was relatively large, we could detect the PCR amplicon in cells targeted with this therapeutic donor. One possible explanation is related to the proliferative advantage of FA-A gene targeted cells over the non-targeted ones. This selective advantage increments the percentage of corrected cells in culture, allowing their detection by PCR.

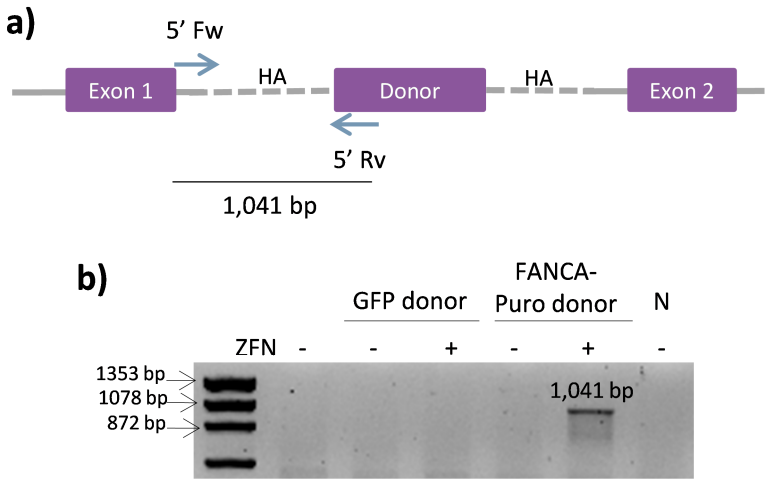


Figure 64: Integration analysis in the 5' end of the AAVS1 locus by PCR of gene targeted FA-55 LCL nucleofected with the FANCA-Puro or GFP donors as plasmid DNA. *a) Schematic representation of gene targeted AAVS1 locus with GFP or FANCA-Puro donors. Arrows represent the primers used to evaluate if*

the integration was site-specific. b) PCR of the bulk population of gene targeted FA-55 LCL cells nucleofected with the FANCA-Puro or GFP donors as plasmid DNA. N: negative control.

The reversion of MMC sensitivity in these cells was also tested. The survival curve showed that cells nucleofected with both FANCA-Puro and ZFNs were resistant to MMC. Surprisingly, cells nucleofected only with the FANCA-Puro donor were also MMC resistant (Figure 65), suggesting that random integration of the donor construct took place in FA-A cells, probably due to the genetic instability of these cells.

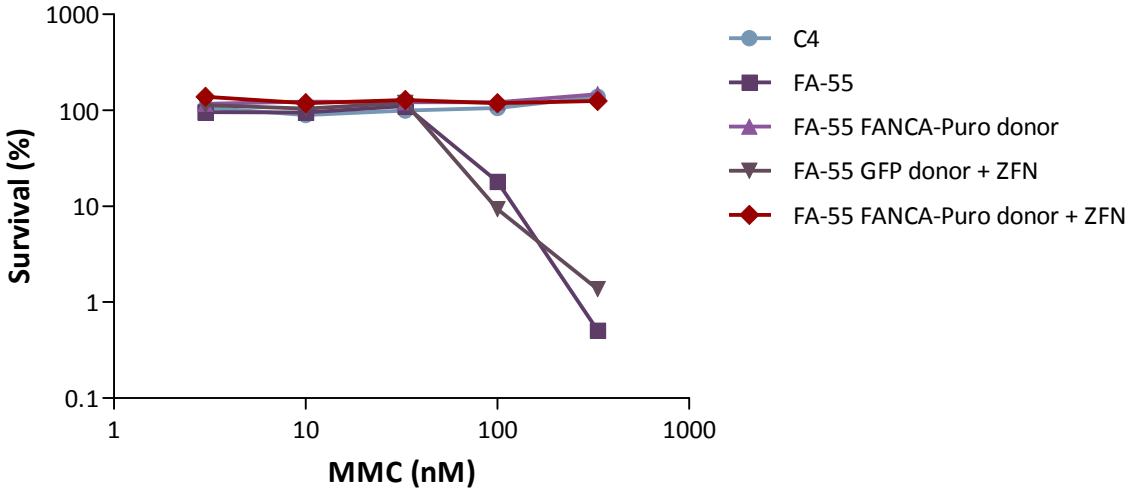


Figure 65: MMC survival curve of FA-55 LCL cell line after gene targeting experiment in which cells were nucleofected with the FANCA Puro or GFP donors as plasmid DNA.

2.2.3. Gene targeting efficiency of the FANCA-Puro and GFP-FANCA donors delivered as IDLV in FA-A LCLs

In subsequent experiments, we investigated whether the specificity of the targeted gene therapy in FA-A LCLs could be improved by the delivery of the donor construct as an IDLV. We used our two therapeutic vectors: FANCA-Puro donor IDLV and GFP-FANCA donor IDLV in the LCL FA-56.

Representative PCR showing the integration of GFP-FANCA donor IDLV in the *AAVS1* in the bulk population of gene targeted FA-56 cells is represented in Figure 66. Only when ZFNs were present the expected band appeared.

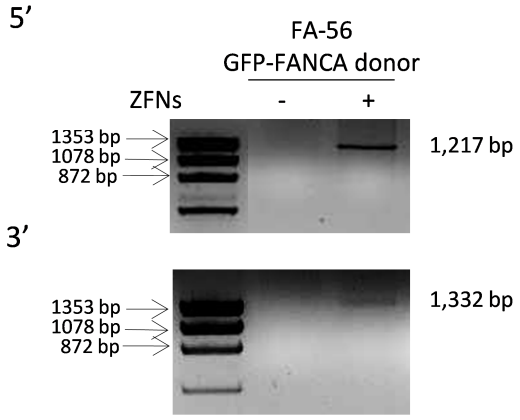


Figure 66: Integration analysis in the 5' and 3' end of the AAVS1 locus by PCR of gene targeted FA-56 LCL nucleofected with the GFP-FANCA donor IDLV.

2.2.3.1. Functional assays

Different functional assays were conducted in these cells to test the phenotypic correction of these cells.

Firstly, we tested the MMC sensitivity of uncorrected and gene edited FA-56 cells. Cells were then cultured with different concentrations of MMC and their survival analyzed at 5 days post treatment. These analyses revealed that there was a reversion in the sensitivity to MMC in FA-56 cells either targeted with the ZFNs and the FANCA-Puro or GFP-FANCA donors IDLV (Figure 67 a, b). In this case the correction of MMC hypersensitivity was only evident when the ZFNs were included. This suggested that the packaging of the FA-donor construct in IDLV facilitated the specific integration, and thus the phenotypic correction of FA-56 cells, with no random integration as it was deduced from the experiments where the donor was used as plasmid DNA (Figure 65).

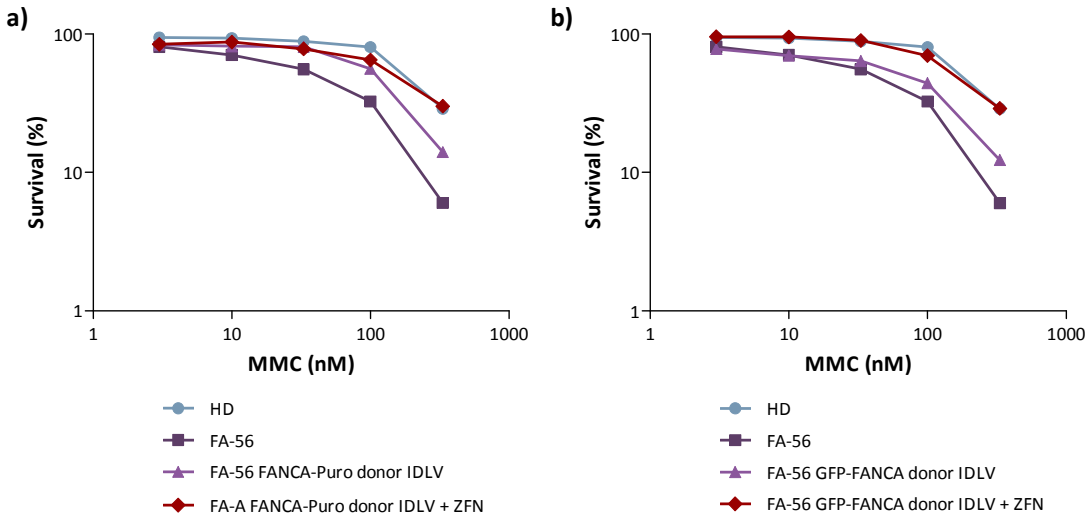


Figure 67: MMC survival curve of FA-56 LCL after gene targeting experiments. *Cells were nucleofected with the FANCA-Puro donor IDLV in panel a, or GFP-FANCA donor IDLV in panel b. HD C3 and FA-56 LCLs were used as controls. MMC resistance was evaluated 5 days after the drug treatment.*

Finally, immunofluorescence studies in FA-56 gene edited cells clearly showed that FANCD2 foci are restored only when gene editing was conducted both with FANCA-Puro and GFP-FANCA donors IDLV together with ZFNs (Figure 68). Even more, the absence of FANCD2 foci when only the donor was used suggest that random integration was not occurring.

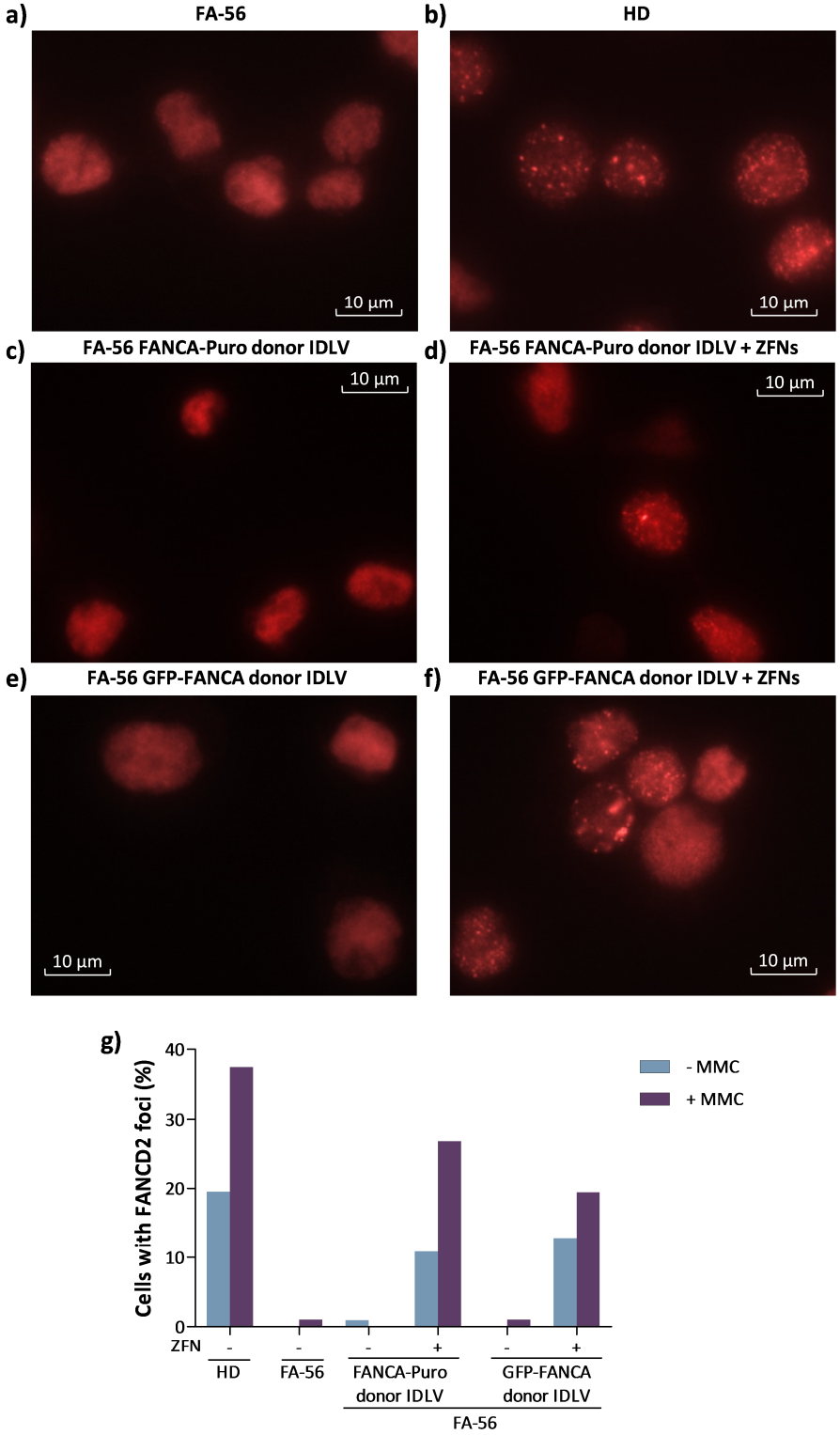


Figure 68: Nuclear FANCD2 foci formation in FA-56 LCLs after gene targeting. *Gene targeted cells were treated or not with MMC and FANCD2 staining was performed (Texas-red). In all instances MMC was*

used to induce DNA damage before FANCD2 immunofluorescence was performed. a) FA-56 LCL. b) HD LCL. c) FA-56 FANCA-Puro donor IDLV. d) FA-56 FANCA-Puro donor IDLV with ZFNs. e) FA-56 GFP-FANCA donor. f) FA-56 GFP-FANCA donor with ZFNs. g) Percentage of cells with FANCD2 foci. HD and FA-56 were used as positive and negative controls respectively and FANCD2 foci were analyzed. Cells with more than 10 foci were counted as positive cells.

2.3. Influence of the PARI inhibition upon the efficacy of gene targeting in healthy donor and FA-A cells

Previous studies have shown that PARI inhibition in FA-D1 cells increases the efficiency of HR leading to a reduction in chromosomal instability[150]. Nevertheless, FA-D1 cells presents a marked defect in HR; while this defect is only mild in FA-A cells[151]. We wanted to test if the consequences of PARI inhibition upon HR were also extensive to FA-A cells and, therefore, if the transient inhibition of this protein could increase the efficiency of HDR-mediated gene targeting. We tested this hypothesis in two different types of FA-A cells: FA-A LCLs and FA-A fibroblasts.

2.3.1. Efficiency of gene editing in lymphoblastic cell lines from FA-A patients after PARI inhibition

First we evaluated the effect of PARI inhibition in a FA-A cell line (FA-55) as compared to a HD LCL (CP1). FA-55 and CP1 were both interfered using PARI iRNAs. Thereafter, the level of PARI inhibition was analyzed by RT-Q-PCR. We showed a reduction in PARI expression of 52.22±0.03% in CP1 and of 68.12±0.08% in FA-55. Since PARI favors the release of RAD51 from the nucleofilament formed during HR; we analyzed the number of cells with RAD51 foci in PARI interfered and in non-interfered cells. As shown in Figure 69 a, the presence of RAD51 foci in HD cells remained constant regardless of the inhibition of PARI or the treatment with MMC. However, the number of cells with RAD51 foci was clearly increased in FA-55 cells treated with the PARI iRNAs both in the absence or presence of MMC, suggesting that the rate of HR in these cells was increased by PARI inhibition (Figure 69 a). To evaluate if the increase in HR due to PARI inhibition induced an increase in the chromosomal stability in FA-A cells, metaphases of non-interfered and PARI interfered FA-A LCLs and HD LCLs were studied (Figure 69 b, c). Our results showed that the interference of PARI in FA-A cells decreased the intrinsic genetic instability of these cells in the absence of DNA damage (- DEB), while did not modify the very low spontaneous genetic instability of HD cells (Figure 69 b, c). After DNA damage with DEB, PARI inhibition did not mediate significant effects, either in HD (which did not show significant aberrant chromosomes) nor in FA-A cells, characterized by a very high number of aberrations (around 60% of cells showed chromosomal aberrations).

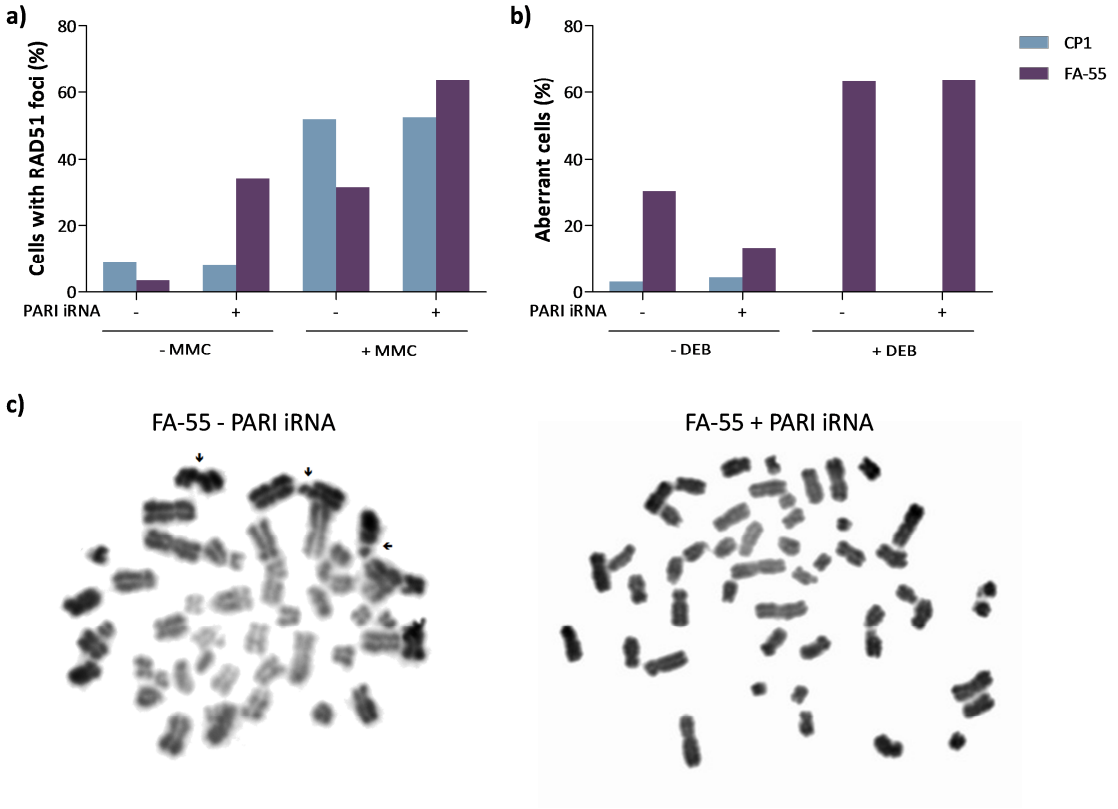


Figure 69: Effect of PARI inhibition in RAD51 foci formation and in the genetic stability of HD and FA-A LCLs. *FA-A and HD LCLs were interfered with PARI iRNAs. a) Analysis of the percentage of cells with nuclear RAD51 foci. Interfered FA-A and HD cells were treated or not with MMC and then fixed and stained with RAD51 antibody. Cells with more than 10 foci per cell were counted as a positive cell. b) Percentage of aberrant cells analyzed in metaphases of interfered HD and FA-A LCLs treated or not with diepoxybutane (DEB). c) Image of different metaphases analyzed in FA-55 samples interfered or not with PARI iRNAs. Arrows indicate chromosomal breakages.*

Once we confirmed that PARI inhibition increased the number of cells with RAD51 foci, suggesting an increase in HR, gene targeting experiments were conducted in PARI-interfered FA-A and HD LCLs, to test if the transient inhibition of PARI correlates with an increased efficiency of gene targeting. Cells were PARI-interfered and one day later transduced with the GFP donor IDLV. Finally, these cells were nucleofected with the ZFNs and the efficiency of gene targeting was measured by flow cytometry. Additionally, the percentage of INDELs was measured by the Surveyor assay. As shown in Figure 70 a, the ratio of INDELs in interfered versus their no-interfered counterparts was comparable in HD and in FA-A cells. The efficiency of gene targeting in these conditions showed that PARI inhibition did not significantly increase the ratio of GFP positive cells either in FA-A or in HD LCLs (Figure 70 b). The high variability in the efficiency of gene targeting between experiments and the relative low number of experiments so far conducted (n=3) did not allow us to obtain significant differences, but only a tendency for increased gene targeting efficiency when PARI was inhibited in FA-A cells.

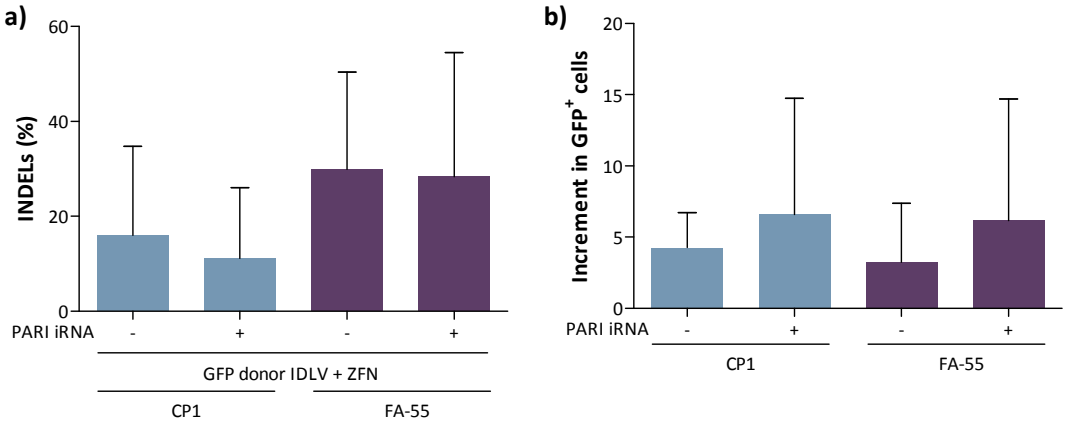


Figure 70: Gene targeting efficiency in PARI-interfered FA-A or HD LCLs. *FA-55 and CP1 cell lines were interfered with PARI iRNAs and then gene targeting experiments were conducted using a GFP donor IDLV. a) Percentage of INDELs measured by Surveyor assay during the different experiments. b) Efficiency of gene targeting represented as the ratio between the percentage of GFP positive cells in cells transduced with the GFP donor IDLV and nucleofected with ZFNs and cells transduced only with the GFP donor IDLV. Bars represent mean and SD. n=3.*

Finally, to verify that the integration of the donor was taking place in a site-specific manner, integration analysis in these samples was carried out by PCR. Our results demonstrated a site-specific integration in these cells only when cells were transduced with the GFP donor IDLV and nucleofected with ZFNs, in both HD and FA-A LCLs (Figure 71).

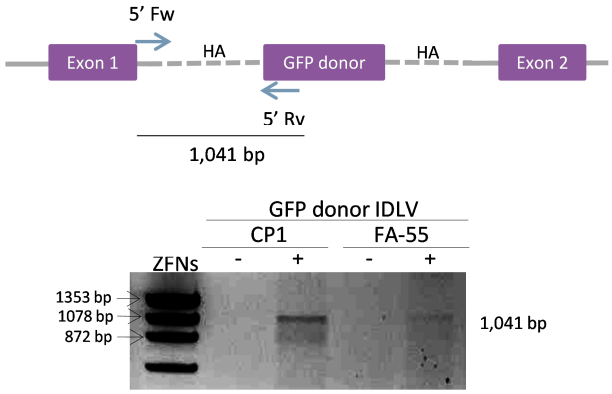


Figure 71: Integration analysis by PCR of PARI-interfered LCLs from a FA-A patient or a HD after gene targeting with a GFP donor IDLV. *Top: Schematic representation of gene targeted AAVS1 locus with the GFP donor. The arrows represent the primers used to evaluate if integration was site-specific. Bottom: Representative PCR of PARI-interfered LCLs from FA-A and HD LCLs.*

2.3.2. Efficiency of gene editing in fibroblasts from FA-A patients after PARI inhibition

Due to the difficulty to transfect LCLs, we also conducted gene targeting experiments in FA-A fibroblasts, trying to clarify the results previously obtained. In these experiments we tested the efficiency of PARI inhibition in fibroblasts from a FA-A patient (FA-123) and from a HD (HFF), in both cases immortalized with hTert. The inhibition of PARI in both types of fibroblasts reached a level of 70% of PARI downregulation (Figure 72 a). One day after the second round of

transduction with the PARI iRNA, cells were transduced with the GFP donor IDLV and one day later nucleofected with ZFNs. As previously observed in LCLs, the gene targeting efficiency in FA-A fibroblasts was not significantly increased when PARI was inhibited, probably due to the high variability in the efficiency obtained in the three experiments that we have conducted so far (Figure 72 b).

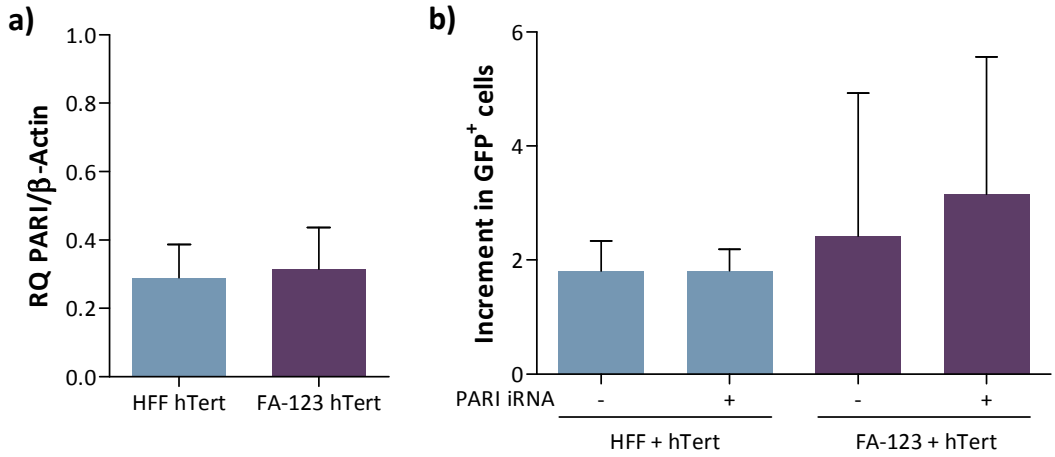


Figure 72: Gene targeting efficiency in PARI-interfered FA-A or HD fibroblasts. *Fibroblasts were interfered with PARI iRNA and then gene targeted using GFP donor IDLV. a) Inhibition of PARI measured by Q-PCR. Bars represent mean with SD. n=6. b) Efficiency of gene targeting represented as ratio of the percentage of GFP positive cells in cells transduced with GFP donor IDLV and nucleofected with ZFNs over cells only transduced with the GFP donor IDLV. Bars represent mean and SD. n=3.*

2.4. Targeted gene therapy in CD34$^+$ cells from FA-A patients

To test the efficiency of our targeted gene therapy approach with the therapeutic FANCA-Puro donor IDLV in hematopoietic stem and progenitor cells from FA-A patients, a small proportion of mPB CD34$^+$ cells that remained in cell collection bags and tubes from the CliniMACS® System used to purify CD34$^+$ cells from FA-A patients included in the FANCOSTEM mobilization trial were used after permission consent of the patients or their parents and approval of the corresponding ethics committees. Cells remaining in the immunoselection bags were prestimulated for 24 hours and then transduced with the GFP donor or the FANCA-Puro donor IDLVs. One day after transduction, cells were washed and nucleofected with 6 µg of the ZFN mRNA. To test the phenotypic correction of the FA CD34$^+$ cells after gene targeting, cells were plated in methylcellulose in the presence or absence of MMC. Numbers of colonies obtained after 14 days in culture were analyzed. In the first two cases (02002 and 01003); cells from patients were transduced with the therapeutic FANCA-Puro donor and the GFP donor IDLV as a control. In both patients an increase in the number of MMC resistant colonies was observed when cells where transduced with the FANCA-Puro donor and nucleofected with the ZFNs (Figure 73 a and b). Since these percentages were higher than the ones expected from our results in HD CD34$^+$ (see data obtained in Figure 41); and to discard that this high efficiency was due to the random integration of the donor vector; in the next two patients (02003 and 02004) the cells were also transduced with the FANCA-Puro donor IDLV in the absence of ZFNs. Cells from patient 02003 presented a percentage of resistant colonies of 4.69% in the

condition transduced with the donor FANCA-Puro and nucleofected with the ZFNs, while no MMC-resistant colonies were obtained in the absence of ZFNs (Figure 73 c). In the last patient, a 10% of MMC resistant colonies were obtained when cells were only transduced with the FANCA-Puro donor IDLV, pointing out that random integration of the donor or an incomplete MMC selection could occur. The percentage of MMC resistant colonies in the condition transduced with the therapeutic vector and nucleofected with ZFNs was 13.16%; suggesting an efficiency of gene targeting of about 3% (Figure 73 d). Overall, our data suggests an increase in the number of MMC resistant colonies when cells were treated with both the FANCA-Puro donor IDLV and the ZFNs, as compared with samples treated either with the GFP donor IDLV or with the FANCA-Puro donor IDLV in the absence of the ZFNs. This data suggests for the first time that gene targeting is occurring in HSPCs from FA-A patients.

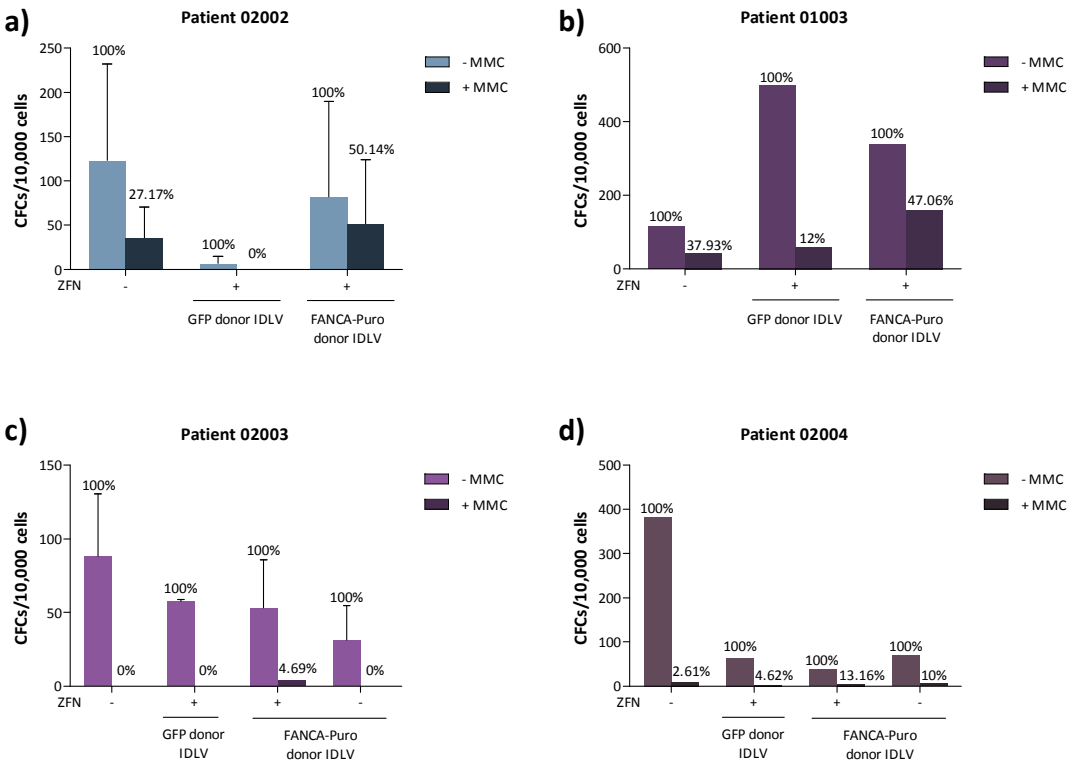


Figure 73: MMC resistance in HPCs from four different FA-A patients after gene editing. *The number of CFCs per 10,000 cells plated is shown both in the presence or absence of MMC using the GFP-donor or FANCA-Puro donor IDLV vectors combined with ZFNs or not. Bars represent mean with SD.*

Integration analysis by PCR of pooled colonies grown in the absence or the presence of MMC was performed to analyze the targeted integration of the FANCA-Puro donor in the *AAVS1* locus. As shown in Figure 74, specific integration in the 5' end of *AAVS1* locus was only observed when cells were cultured in the presence of MMC, demonstrating for the first time that targeted gene therapy in FA-A $CD34^+$ cells is feasible, although at lower efficiencies as compared with those obtained in HD $CD34^+$ cells targeted with smaller non-therapeutic donors (see data from experiments shown in Figure 41).

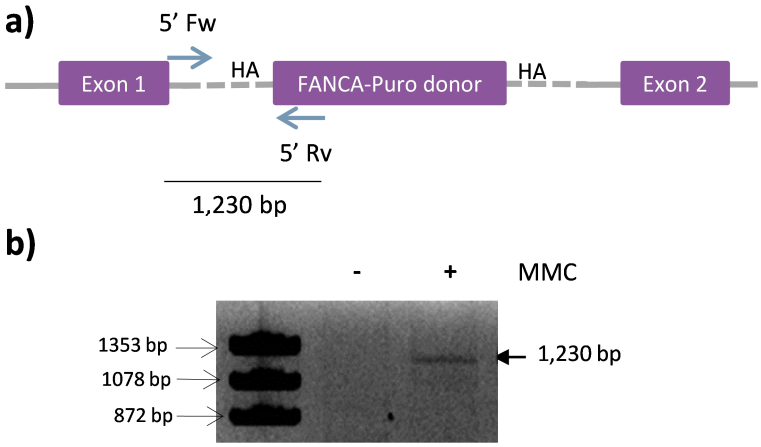


Figure 74: Integration analysis by PCR of hematopoietic colonies obtained from patient 02002 CD34$^+$ cells after gene targeting using FANCA-Puro donor IDLV. *a) Schematic representation of gene targeted AAVS1 locus with FANCA-Puro donor. Arrows represent the primers used to evaluate the specific integration of FANCA-Puro donor in the 5' end of the AAVS1 site. b) PCR of a pool of colonies from gene edited FA-A CD34$^+$ cells treated or not with MMC.*

DISCUSSION

1. General aspects of gene editing applied to the therapy of inherited monogenic diseases

Clinical trials conducted with gamma-retroviral vectors (γ-RVs) presented some adverse effects derived from the insertion of transgenes in transcription start sites and upregulation of neighboring genes. Some of these insertions mediated the upregulation of proto-oncogenes inducing leukemia in clinical trials of X-linked severe combined immunodeficiency (SCID-X1)[26,27], Wiskott-Aldrich syndrome (WAS)[17] and chronic granulomatous disease (CGD)[15] patients. In order to increase the safety of hematopoietic stem cell (HSC) gene therapy (GT), self-inactivating (SIN) vectors, with a very limited trans-activation potential, due to inactivation of the viral LTRs, have been developed. Moreover, the use of SIN-lentiviral vectors (LVs), which additionally present a safer integration profile than γ-RVs, have also contributed to increase the safety of HSC GT. Significantly, current GT trials with SIN-LVs and SIN-γ-RVs have shown clinical efficacy in the absence of severe side effects in several monogenic diseases such as WAS[18], metachromatic leukodystrophy (MLD)[23], SCID-X1[12-14], adrenoleukodystrophy (ALD)[22], β-thalassemia[21] and sickle cell anemia[303,304].

Very recently, gene editing has appeared as a new and even safer GT platform that facilitates specific integration of the transgenes in desired genomic loci. However, gene editing is still promising and should be improved in order to be extensively applied in the clinics. The three key parameters of gene editing are specificity, efficiency and fidelity; being the combination of all of them what would facilitate their use in the clinics based on the optimized development of designed nucleases and donors and also an efficient approach for the gene transfer.

The specificity in the gene editing is defined by the probability of integrating the selected transgene in the certain site, without affecting any other sequence of the genome. It can vary depending on the selected reagents and experimental conditions. For example, there are evidences that the short-term exposure of cells to nucleases reduces their off-target activity and their related cytotoxicity[127,130]. Acute cytotoxic effects may possibly result from ATM or ATR-dependent protein kinase signaling pathway activation, leading to cell cycle arrest and apoptosis[305]. This fact, together with the involvement of off-target DSBs in the generation of either, unintended mutations, illegitimate recombination events or translocations, makes nuclease specificity an important parameter in gene editing. Therefore, the development of designed nucleases is often associated with considerable efforts directed towards maximizing their specificity; for example by the generation of obligate heterodimeric variants in ZFNs and TALENs[102,103]. Some genome-modifying events may affect the specificity of gene editing for example by the targeting of the donor DNA in breakage-prone fragile sites, or in sites with DSBs generated by DNA metabolites or environmental mutagens[306].

The second key parameter, the efficiency of gene targeting is mainly mediated by delivery of genome editing tools into the target cells, which highly depends on the target cell type. For example, in the case of the HSCs, difficulties related to their transfection imply that the

delivery of nucleases and donors rely on other approaches such as transduction with viral vectors or electroporation[117]. However, viral vectors are sometimes difficult to produce due to the size of the donor transgenes. On the other hand, electroporation although nowadays is not as toxic as it used to be, still induces apoptosis and cell death in many different cells types. Due to all these reasons, an improvement of these technologies is necessary in order to efficiently conduct gene targeting.

Finally, the fidelity of gene editing is based on the efficiency of the process to restore the sequence of the edited locus, without generating not intentioned mutations on it. This will reduce collateral adverse genome-modifying events derived from the targeting process. In this regard, homology-directed repair (HDR) strategies mediated by nucleases are multifaceted because their ultimate efficacy depends not only on the nuclease specificity but also on the fidelity with which the exogenous DNA is inserted at the targeted site. Fidelity is also compromised by the insertion of concatemeric sequences generated through non-homologous recombination processes, such as NHEJ. Although targeted, sometimes the integrated sequences can introduce vehicle-derived backbone fragments, such as viral or bacterial sequences, into the cellular DNA. Moreover, in nuclease-exposed populations, a substantial fraction of targeted alleles will undergo gene disruption instead of homology-directed gene targeting. These stems from the fact that NHEJ occurs throughout the cell cycle and competes with HR for DSB repair[307]. The use of nickase variants of nucleases, which generate single strand breaks in the genome instead of DSB, has shown to trigger the repair of these lesions through HDR rather than NHEJ[308,309]. Finally, it has also been proven that, for the gene correction of recurrent point mutations, donor sequences consisting in single stranded oligonucleotides offer a simpler and faster alternative to plasmid and viral vector donor DNAs[121].

In the early stages of gene targeting, the main focus of the research was the efficiency of the process. However, once the efficiency is reached, improvements in the specificity of the nucleases and in the fidelity of the approach are also required.

Gene editing is a growing and a trending topic field in gene therapy. Most of the leader gene therapy teams are working in therapeutic and applied aspects of gene editing, and just this year, more than 190 papers have been published about this topic. From the clinical point of view, the most advanced nucleases are the ZFNs, probably because they were the first developed nucleases and also due to their high efficiency in hematopoietic cells. In this respect, nowadays there are seven different clinical trials to treat HIV-1 patients using ZFNs that target *CCR5*, a co-receptor used by the virus to enter into T cells. These trials are based on the disruption of this gene using specific ZFNs in $CD4^+$ cells and also in hematopoietic stem and progenitor cells (HSPCs)[62,110,111]. Furthermore, two preclinical studies to treat two different monogenic disorders (X-SCID[113] and sickle cell anemia[115]) have been already published based on a ZFN-mediated knock-in strategy. There is also a very advance study in which adeno-associated vectors (AAVs) were used to carry ZFNs and donors to treat *in vivo* hemophilic mice using a knock-in strategy[134,310].

The final aim of our study is to develop a targeted GT approach in $CD34^+$ cells from Fanconi anemia A (FA-A) patients. In terms of specificity we have used ZFNs specifically targeting the

AAVS1 locus, whose safety and efficiency has been already demonstrated[65,69]. Moreover, we have used ZFNs in which the *Fok*I nuclease presents mutations that obligate it to work as a heterodimer[102,103]. Aiming to increase the efficiency of gene targeting, we have investigated the use of gold nanoparticles (AuNPs) to increase the transduction efficiency of lentiviral vectors, including integrase-defective lentiviral vectors (IDLVs). Potential increments in transduction efficiency could be associated with increases in the percentage of targeted cells, since more cells would incorporate the donor construct. Finally, in terms of fidelity we have investigated the potential increment of homology directed repair (HDR) in FA-A by the inhibition of an antirecombinase protein called PARI.

2. Gene targeting with control vectors in healthy donor hematopoietic cells

As it has been mentioned, gene editing strategies represent the best alternative to conventional GT, since the insertion of the gene of interest in the cell genome is controlled, avoiding the potential risk of insertional mutagenesis[2]. The application of this approach in the clinics appeared really improbable only a few years ago, due to the extremely low frequency of HR. However, the development of different types of nucleases generating DSBs in a specific site of the genome, has increased the frequency of HR up to 10,000 times[311], getting the application of gene targeting closer to the clinics. Therefore, genome editing technologies based on programmable nucleases are opening up the possibility of achieving therapeutic genome editing in diseased cells, resulting in the removal[312] or correction of deleterious mutations, or the insertion of wild type genes. Designed nucleases enable the precise genome editing by introducing DNA DSBs at specific genomic loci. DSBs subsequently recruit the endogenous machinery of the cells to repair by NHEJ or HR the DSBs, facilitating in the case of HDR genome editing. The DNA repair mechanism selected by the cell depends on its cycling state, and also on the presence of a repair template: the donor. During the present thesis we have established a robust protocol of gene targeting for different types of hematopoietic cells, mainly lymphoblastic cell lines (LCLs) and hematopoietic stem and progenitor cells (HSPCs) ($CD34^+$ cells) using ZFNs as a tool to facilitate the targeting in the *AAVS1* safe harbor locus.

In our case we selected the *AAVS1* safe harbor strategy to accomplish our gene targeting experiments for several reasons: *AAVS1* locus has been deeply characterized. It consistently allowed higher transgene expression as compared with another safe harbor locus, such as the *CCR5,* in several human cell types including $CD34^+$ cells[113], when different constructs were used. Additionally, the promoter of the *PPP1R12C* gene, located in the *AAVS1* locus, facilitated in most instances the expression of the transgenes since this gene is constitutively transcribed in most cell types. Moreover, the chromatin in this locus presents an open conformation, which facilitates the incorporation of the transgenes and the expression driven from the exogenous promoters inserted together with the transgene. Even more, transgene insertion in *AAVS1* locus do not upregulate the targeted and flanking genes[65], and finally, and more importantly, the interruption of the *AAVS1* locus does not produce any pathological effect[68].

These observations let us to predict that transgene insertion should have no detrimental consequences in targeted HSPCs.

Gene targeting mediated by HDR could be achieved in three different ways: gene correction, in which the mutated sequence is replaced by the original one; knock-in, in which a partial cDNA of the gene of interest is introduced in its endogenous locus; and safe harbor, in which the whole expression cassette is inserted in a safe site in the genome. We considered that safe harbor approaches may constitute an advantage in FA, since 19 different complementation groups have been described so far. This implies that other HDR-mediating gene targeting strategies, such as knock-in, would require the development and testing of 19 different nucleases and donors. Furthermore, gene correction, although it could be possible in FA-A patients from the intron 2 of the protein (the majority of mutations are allocated after the second intron of the gene[59]), it would require the generation and deeply characterization of new nucleases and donor substrates. Our selected safe harbor strategy presents the additional advantage that highly characterized ZFNs designed for the *AAVS1* site have been already generated and used in different studies[65,69,75]. In contrast, the main disadvantage of the safe harbor strategy is the loss of the endogenous regulation of the inserted gene as it is not located in its original locus. However, this potential limitation could be solved, at least in part, using the endogenous promoter of the therapeutic gene if it is known, as is the case of *FANCA*[313]. Using this kind of strategy and only changing the transgene and maybe the promoter, any monogenic disease could be in principle treated by this approach. Nevertheless, a deep study of the regulation of the gene of interest to evaluate potential problems derived from the down- or up-regulation of the therapeutic protein in the cell of interest are necessary. In the case of FA, a previous study from our laboratory has shown that FANCA expression mediated by ubiquitous PGK promoter allows the correction of the phenotype in hematopoietic cells from FA-A patients reaching FANCA expression levels similar to healthy donors (HD). Even more, this study also showed that levels of FANCA expression are not a critical point since both, low levels of expression directed by VAV promoter, and supra-physiological levels directed by SFFV promoter corrected the FA phenotype in hematopoietic cells[290]. Moreover, experiments conducted in our laboratory overexpressing h*FANCA* in healthy donor cells, did not induce any toxicity[290].

Zinc finger nucleases are the only nucleases that have been approved so far in clinical trials to treat HIV-1 patients by knocking out the *CCR5* receptor, which is a co-receptor used by the virus to enter into the cells[110]. In this respect, it has been demonstrated that the disruption of this gene in $CD4^+$ cells did not mediate any deleterious effect[62]. It has also been demonstrated that the use of ZFNs to target the *CCR5* in $CD34^+$ cells is also feasible, and could be a long term treatment for patients infected with HIV-1. In that study the nucleofection of the ZFNs as plasmid DNA induced a 17% of disruption of *CCR5*, maintaining the differentiation potential of edited cells to any lineage and their repopulation capacity in immunodeficient animals[111]. However, considering that the *AAVS1* locus supports higher transgene expression than *CCR5*, we selected ZFNs targeting the *AAVS1* locus as the most appropriate tool to conduct our gene targeting studies.

The target cell type is a major issue in gene targeting approaches. In most of the hematopoietic diseases the target cell of interest is the HSCs, since the correction of this

population of cells will lead to the generation of corrected cells in the different hematopoietic lineages. However, in contrast to other hematopoietic diseases in which a specific lineage is affected, in the case of FA the hematopoietic stem cell is also affected, leading to a defective generation of mature cells and finally to the BM failure. This complicates gene targeting approaches in FA compared with other hematopoietic diseases.

Targeting of HSCs with nucleases is really challenging. There are different strategies to introduce the ZFNs in hematopoietic cells: the cells could be transduced with non-integrative viral vectors such as IDLVs[112] or AdVs[314] carrying the nucleases, or they could be nucleofected with the ZFNs as plasmid DNA[111], mRNA[113,115] or protein[315]. Although the transduction with these kind of viral vectors would not mediate their integration in the genome of the hematopoietic cell, their effects could remain for some time in these type of slowly proliferating cells leading to toxicity[316]. On the other hand, nucleofection has been associated with a relatively high toxicity for $CD34^+$, leading to the loss of engraftment potential, although more recently less toxic approaches have been described[113,115]. IDLVs and nucleofection using plasmid DNA has the problem, although it is improbable, of random integration of the nuclease in the genome leading genomic instability. Moreover, the nucleofection of DNA is always more toxic than nucleofection of mRNA or proteins, due to the activation of cellular immunity mechanisms. The delivery of ZFNs as mRNA or protein ensures a rapid burst of expression and also avoids the possible insertional risk. However, the *in vitro* synthesis of proteins is more technically difficult than the *in vitro* synthesis of mRNA, so we decided to select the delivery of the ZFNs as mRNA in our experiments. Moreover, ZFNs used as mRNA showed to be a good method to conduct gene targeting in $CD34^+$ cells[113,115]. We tested two different pairs of ZFNs: one of them without any regulatory region, and the other with the regulatory regions of the tobacco etch virus and human *β-globin,* based on previous studies shown that modifications in RNA during the synthesis of the nucleases increment its activity[317]. In our hands, the presence of these regulatory regions diminished the cutting efficiency of the ZFNs (Figure 29), and therefore reduced the gene targeting in LCLs (Figure 30 b), and also completely abolished the gene targeting in hematopoietic progenitors, (Figure 38).

One of the most important parameters to analyze when a new nuclease is developed is its off-target activity that can lead to the unspecific integration of the donor cassette, potentially inducing genetic instability. To minimize the off-target activity, we selected the *AAVS1* ZFNs developed by Sangamo Bioscience which have been deeply characterized. Although one published report showed that *AAVS1* TALENs were more specific than ZFNs, only very low off-target effects were detected[66]. These ZFNs were firstly studied in the work of Hockemeyer and coworkers[69], in which they demonstrated by Southern blot analysis that gene targeting in an embryonic stem cell line only presented the site-specific integration. Moreover, in the work of Rio et al[75] we also demonstrated by Southern blot that the integration mediated by these ZFNs was also site-specific. In the study of Lombardo and coworkers[65], the authors analyzed how the integration of a promoter and a transgene affected the expression and regulation of neighboring genes in different cell types, showing that the use of the PGK promoter together with a transgene did not alter the expression of neighboring genes. These ZFNs were also used in the study of Genovese and coworkers[113], in which they efficiently targeted $CD34^+$ cells using a GFP donor and they observed by Southern blot analysis of iPSCs derived from GFP^+ sorted gene targeted cells, that the integration in this locus was site-specific.

We also tested different ways to deliver the donor to lymphoblastic cell lines. First we used the donor as plasmid DNA. This is easy to produce and also cheap, but presents a potential risk of random integration. Although plasmid donors have not been used in HSPCs, probably due to the toxicity and difficulty to deliver this molecule in this kind of cells, its efficiency has been proven in other set of cell types, such as K562[70,97,118], $CD4^+$ T cells[118] and human embryonic stem cells (hESCs)[69] accomplishing efficiencies between 7.7-18%, 5.3% and 50% of hESCs selected clones respectively.

Then, we used the donor as IDLV. IDLVs have been already tested in LCLs[65,112], reaching efficiencies between 5.4-12% and in $CD34^+$ cells[113,115] with efficiencies between 6-18%. The main limitation of IDLV is their production since IDLV vector production often results in lower titers as compared with ICLVs, limiting the transduction of primary cells, and therefore reducing the gene targeting efficiency. Moreover, the variability observed between IDLV productions is higher than in ICLV productions. Using donor as DNA we could achieve gene targeting in LCLs, but the frequencies that we obtained were lower than the ones obtained with IDLV (around 3% using DNA as donor compared with 10% using IDLVs) (Figure 32, Figure 33). The levels of gene targeting that we obtained using IDLV as donor are comparable with the ones obtained by Lombardo and coworkers using the donor and the ZFN as IDLVs[65]. Taking this into account, we selected the donor as IDLV for our gene targeting experiments in $CD34^+$ cells. Consistent with our studies, the only two papers published in gene editing in $CD34^+$ cells also used IDLV as the method to deliver the donor construct[113,115].

When gene targeting is performed, it is important to verify that the integration is site-specific. In order to do that, different approaches can be performed. Sequencing of the entire genome of single clones is time consuming and expensive, but is the most accurate technique to detect the overall targeting events that are occurring in each cell. Another possibility is to conduct only the sequencing of the targeted sequence. An alternative strategy that can be used and that gives information of the targeting events in the entire genome is the Southern blot analysis. However, the amount of sample required to perform this technique is not always possible to collect, especially when the target population consist on primary cells in which the expansion is limited as it is the case of HSPCs. Finally, the most frequently used approach to verify if the integration is site-specific is the PCR, using primers that hybridize outside and inside of the donor cassette using both integration ends, the 5' and the 3'. Using this technique we have only detected on-target events in the desired locus. However, we cannot be sure if some concatemeric insertions were generated, or if there are a few off target events. Although PCRs interrogating both sides of the targeted locus constitute a routine widely used analysis, it presents a limit of detection when bulk populations are used. We investigated the PCR detection limit in a sample in which gene targeting was confirmed both by flow cytometry and also by PCR (Figure 35). We estimated that percentages of integration below 2.25% in a bulk population induced with our GFP donor IDLV were not detected by PCR. One possibility to overcome this problem, in the case in which the gene targeting is performed with some reporter genes such as *GFP*, is to sort GFP positive cells in order to enrich the gene targeted cell population as it has been done in previous reports[113].

In order to conduct gene targeting in $CD34^+$ cells, we pre-stimulated the cells for two days to promote the cycling of the primitive HSCs, most of which are quiescent in bone marrow (BM), umbilical cord blood (UCB) or mobilized peripheral blood (mPB)[318]. Cycling is necessary to activate the HR in these cells. Otherwise DSBs generated by the ZFN would only be repaired by NHEJ, since HR is only active in the G2/S phases of cell cycle. The increased HDR-mediated targeting in pre-stimulated HSCs was also demonstrated in the study of Genovese and coworkers[113] in which they observed a significant increase in the percentage of GFP^+ cells when $CD34^+$ cells were pre-stimulated for two days before conducting the gene targeting approach. Although these authors added some molecules to prevent the differentiation of the HSCs, such as aryl hydrocarbon receptor antagonist (SR1) and/or 16,16-dimethyl-prostaglandin E2, we saw that we could achieve slightly increased percentages of GFP^+ cells in the primitive HSCs ($CD34^+$ $CD133^+$ $CD90^+$) in the absence of these compounds (4.5% of GFP^+ cells compared with 3% measured at day 7 of culture[113]) (Figure 42 a) without affecting its engraftment and repopulating capacity (Figure 46). In agreement with the study of Genovese et al[113], the lowest gene target efficiency is observed in the most primitive subset (Figure 42). Eight different experiments were carried out in $CD34^+$ cells using different samples of thawed UCB, which revealed the reproducibility of the targeted GT (Table 12). On average we obtained 10.42% GFP positive cells (Figure 41), which reflected a gene targeting efficiency of 9.43%, considering that 90.48% of the GFP positive colonies had the specific integration of the donor in the *AAVS1* site. The efficacy of targeting achieved in our experiments could be enough to obtain a clinical benefit in FA patients, as it has been demonstrated that the spontaneous reversion of only one progenitor cell can reconstitute the hematopoiesis of the patient due to the proliferative advantage that corrected cells present over non-corrected ones[5]. However, our gene targeting efficiencies are lower than the ones obtained in a more recent study (18% in $CD34^+$)[115]. Nevertheless, in this study they targeted the mutated locus using a therapeutic donor IDLV and following a knock-in strategy in comparison with our safe harbor strategy using GFP donor IDLV which was also investigated in the study of Genovese and coworkers[113]. Moreover, the efficiency of gene targeting in Hoban study is measured by Q-PCR assay, and we measured this efficiency by flow cytometry as Genovese et al[113].

Our results showed that the increment of gene targeting efficiency correlates with an increase in the efficiency of ZFN to generate DSBs (measured by %NHEJ), suggesting that the donor is not a limiting factor for gene editing in our approach.

Since our final goal is to conduct gene editing in HSPCs from FA-A patients and to demonstrate their capacity to restore their hematopoiesis in the long term; we also evaluated if gene targeted HSPCs presents multilineage repopulating potential. To test their repopulating capacity, gene edited $CD34^+$ cells were transplanted in sublethally irradiated immunodeficient mice. For this purpose we selected NSG-SGM3 mice strain, which constitutively express three different human cytokines: human interleukin-3 (IL-3), granulocyte/macrophage-stimulating factor (GM-CSF) and steel factor (SF). This mouse strain was developed to increase the engraftment potential of acute myeloid leukemia (AML) cells, a process which also depends on the continuous cytokine stimulation of these cells[319]. As shown in Figure 46 a, we obtained a stable engraftment of human cells in all transplanted mice at 90 dpt, with a median of human engraftment of 17.1%. Furthermore, 83.3% of the transplanted mice presented GFP positive engraftment with a median of 11.9% of GFP^+ cells in the human engraftment (mean: 14±12%)

(Figure 46 b). This value is very similar to the gene targeting efficiency detected in $CD34^+$ cells *in vitro* before the transplantation (10.42±4.91%) suggesting the efficient targeting of the HSPCs and that targeted HSPCs do not present any engraftment disadvantage in comparison with non-targeted ones. These results are in contrast to previous studies, in which the percentage of GFP^+ cells decreased after transplant[113,115]. This could be due to differences in the strains of immunodeficient animals used, as in these studies the authors used the NSG mouse strain, while we used the NSG-SGM3 strain. Since we are directing our gene targeting strategy to the HSPCs, capable to differentiate towards the different hematopoietic lineages, we also confirmed the presence of GFP positive cells in all different subpopulations present in the BM (Figure 47), pointing out that we have targeted HSPCs and that gene editing did not interfere with the multilineage potential of these cells. Moreover, the PCR analysis of the site-specific integration in GFP positive colonies obtained from the BM of transplanted recipients showed that the integration was maintained over time in the HSPCs (Figure 48).

Strikingly, when we used the donor in which the *FANCA* gene expression is driven by the PGK promoter and the *GFP* by the endogenous promoter of the *PPP1R12C*[75] (Figure 16), we did not observe GFP expression in gene-edited LCLs (Figure 36) or in $CD34^+$ cells (Figure 49 a). However, we were able to confirm the targeted integration of this cassette in the *AAVS1* locus by PCR analyses in both cell types (Figure 37, Figure 49 b). In contrast, in a previously published study, GFP expression from *PPP1R12C* promoter was detected in a B lymphoblastic cell line after gene targeting[65], and we were also able to detect GFP expression of this donor cassette in fibroblasts from healthy donors and FA patients[75]. It has been proven that viral titers decreased semi-logarithmically with increasing vector lengths[320] and therefore the transduction efficiency reached with bigger vectors such as our GFP-FANCA donor (14.6 kb) was lower compared to the one obtained using smaller vectors such as GFP donor (9.7 kb). This lower transduction efficiency reduces therefore the gene targeting and this could impede us to detect any GFP signal by flow cytometry in hematopoietic cells, which are cells more difficult to target than fibroblasts.

However, as we were able to detect specific integration of this GFP-FANCA donor in pools of colonies positive for GFP presence by Q-PCR (Figure 49 b), NSG-SGM3 mice were transplanted to evaluate the engraftment and multilineage reconstitution capacity of gene edited cells targeted using GFP-FANCA donor. Our results demonstrated that integration of h*FANCA* cassette in the *AAVS1* locus did not affect the engraftment capacity of these cells (Figure 50). Moreover, we still detected GFP engraftment by Q-PCR and this engraftment could also lead to all the different hematopoietic subpopulations (Figure 51). Additionally, the human engraftment of these cells was similar to the one obtained using the GFP donor (Figure 46 a). Therefore, despite the big size of the donor therapeutic construct, in comparison with previously used donors[113,115], we were able to target $CD34^+$ cells, and this population of targeted cells showed *in vivo* reconstitution capacity in immunodeficient mice.

Since one possible approach to increase the efficiency of gene targeting rely on the improvement in the delivery of the donor and the nuclease, we investigated the efficiency of combining AuNPs with the IDLVs as a potential approach to improve the efficiency of gene targeting. Gold nanoparticles (AuNPs) have been used as carriers to introduce different molecules in many cell types and organisms, both *in vitro* and also *in vivo*. These particles have

been combined with siRNAs by electrostatic charges (AuNP is positively charged and the siRNAs present negative charge)[145] or covalently conjugated with them obtaining good results of gene silencing in different models, from cells to complex organisms such as mice[143,144]. Here, we have demonstrated for the first time that the combination of AuNPs with ICLVs and IDLVs doubles the transduction efficiency of both types of vectors in HSPCs *in vitro* (Figure 53). Moreover, we have proven that AuNP-mediated transduction of LVs in $CD34^+$ increases its transduction efficacy as revealed by *in vitro* assays (Figure 53) and also in short-term repopulating experiments (Figure 54 b). This effect was even more pronounced when lower MOIs were used, probably because with higher MOIs maximum transduction efficiency is reached in all circumstances. In spite of these data, the application of AuNPs to our gene targeting protocol resulted in high toxicity. This is probably due to the nucleofection step used to introduce the ZFN mRNAs in the cells to counteract this effect. One possibility would be to deliver the nucleases and also the donor as IDLV, in combination with the nanoparticles. The delivery of both as IDLV has been already proven in $CD34^+$ cells with very modest efficiency[112], but the use of gold nanoparticles could increase their efficiency maintaining a low toxicity. Recent studies have also demonstrated that the addition of some drugs such as rapamycin[139,140] increases the transduction efficiency of ICLVs and IDLVs in human and mouse HSCs without disturbing the main characteristics of these cells and decreasing the timing of culture of these stem cells. However, when we tested the effect of rapamycin in gene targeting experiments, although we detected an increment in the transduction efficiency, a negative effect in the gene targeting efficiency was observed (data not shown); probably related with other effects induced by rapamycin on the cell, such as inhibition of cell cycle progression[321].

3. Gene targeting in Fanconi anemia hematopoietic cells

Monogenic disorders that affect the hematopoietic system are one of the main targets of conventional gene therapy due to the accessibility of the hematopoietic stem and progenitor cells (HSPCs) and the improvements in hematopoietic stem cell manipulation and transplantation (HSCT). Advances in the development of integrative vectors such as lentiviral vectors (LVs) with a safer integration profile than RVs[24], the development of self-inactivating (SIN) vectors lacking the strong promoter activity of the LTRs, and the use of weaker or cell lineage promoters, has increased the safety of gene therapy (GT), as it has been proved in different successful clinical trials, such as the SCID-X1 clinical trial with SIN-γ-RV[13] or WAS clinical trial with SIN-LVs[18].

In spite of the advances of conventional GT, the site of insertion of the transgene in these conventional GT approaches is not controlled. Gene editing could overcome this limitation, and for that reason many efforts are nowadays focused on the treatment of monogenic disorders, such as FA, using gene editing approaches.

Fanconi anemia cells present a clear advantage to perform gene targeting, which derive from the proliferative advantage of corrected cells over the non-corrected ones. However, the main disadvantage of FA-GT is that the main affected cells are the HSCs, which limits the number of

cells to conduct gene editing. Additionally, the frequency of homologous recombination in FA cells could be low due to their DNA repair defects. In this study, we first tested if gene targeting is feasible in FA-A hematopoietic cells. It has been described that mutations in proteins of the FA core complex result in only mild HR defects in mammalian cells[228] except for HR repair of cross links coupled to replication[232]. Although some FA proteins such as FANCD1/BRCA2 clearly promote HR, members of the core complex are not essential components of the HR machinery[151]. On the other hand, FA cells have levels of NHEJ that are comparable with those of complemented cells[228]. A previous study from our laboratory has also shown the feasibility to conduct gene editing in fibroblast from FA-A patients[75]. However, there is no study showing the efficiency of gene editing in hematopoietic cells from FA patients, which are the preferential target to conduct gene editing, as defects in these cells are responsible of one of the main cause of death in these patients, the bone marrow failure. To clarify if HDR-mediated GT can be conducted in hematopoietic cells from FA patients we performed three independent experiments in four different LCLs from FA-A patients, either uncorrected or corrected with a therapeutic LV, to compare the respective gene targeting efficiency (Figure 59). Our results showed that the correction of the FA pathway in FA-A LCLs mildly increased the efficiency but was not essential for HDR, confirming for the first time the feasibility to conduct gene editing in hematopoietic cells from FA-A patients.

Working with a plasmid donor we have observed that the off-target integrations in FA-A LCLs are more frequent when compared with healthy donor cells (Figure 63). This observation can be explained by the fact that increasing spontaneous DSBs in FA cells can lead to higher increase in random integration. This, together with the fact that cells which incorporate the *FANCA* transgene, either in the *AAVS1* site or randomly, present a proliferative advantage could lead to a population of selected cells with a high proportion of cells harboring the *FANCA* gene in off-target sites. However, in contrast with the high percentage of off-targets observed with plasmid DNA, the use of IDLV vectors markedly improved the targeted integration, and thus the safety of this process. In this case the proliferation advantage of cells that incorporate randomly the donor is an adverse event, as the donor cassette incorporates an ubiquitous promoter and therefore is expressed in any place in which it integrates. One possible solution could be the use of a donor construct with a splicing acceptor site such as E2A without promoter. This direct the expression of the transgene from the endogenous promoter of the *AAVS1* locus, but in this case the level of the protein expression should be evaluated, as we do not know if it could be high enough to correct the FA phenotype.

During the present thesis we have developed a new therapeutic vector (FANCA-Puro donor) because of the difficulties to detect the reporter gene after targeting of the GFP-FANCA donor[75] in the *AAVS1* site of hematopoietic cells (Figure 36 and Figure 49). This new therapeutic donor vector contains the puromycin resistance gene, instead of a fluorescence reporter gene, that would allow us to select targeted HSCs in short periods of time (24 hours)[69], to prevent the *in vitro* differentiation of HSCs. This antibiotic resistance gene is also under the control of the PGK promoter, as it is placed downstream the *FANCA* gene and preceded by an E2A splicing acceptor site. This is in contrast to the *GFP* gene in the GFP-FANCA donor, which expression was driven by the endogenous promoter of the *PPP1R12C* after HDR in the first intron of this locus. To assure the functionality of the new vector we tested it as an ICLV in FA-A LCLs, showing reversion of the classical FA phenotype: reversion of MMC hyper-

sensitivity (Figure 60 a), as well as the re-expression of the FANCA protein (Figure 61) and finally the restoration of the FANCD2 nuclear foci formation after DNA damage (Figure 62). All this data constitutes a proof that the FA/BRCA pathway was reestablished after transduction with this therapeutic vector. In addition to these results we also noted that transduced cells became resistant to puromycin confirming that the selection gene was also working (Figure 60 b).

It is known that GT efficiency depends on transduction efficiency which also depends on the viral titer, a parameter that is related to the size of the transfer vector. Larger vectors lead to a reduction in the titer, and consequently in the transduction efficiency[320]. Taking this into account, the efficient transduction of hematopoietic cells with large vectors, such as the ones carrying the *GFP* and *FANCA* genes (GFP-FANCA vector: 10.7 kb) and the *FANCA* and puromycin resistant genes (FANCA-Puro vector: 10.1 kb), is challenging, since previous studies have used much smaller constructs[113,115]. Interestingly, our results show for the first time that gene targeting can occur with donor vectors such as GFP-FANCA and FANCA-Puro donors in hematopoietic cells from HD (Figure 49) and also from FA-A patients (Figure 64, Figure 66).

As observed after transduction with the GFP-FANCA and FANCA-Puro ICLVs, our gene targeting experiments demonstrated the phenotypic correction of FA deficient hematopoietic cells using our therapeutic donor vectors (Figure 67, Figure 68). This correction restored the FA/BRCA pathway measured by the FANCD2 foci formation (Figure 68), a process that is essential for the repair the ICLs. Members of FA core complex are the ones that allow the monoubiquitation and recruitment of FANCD2-FANCI complex to the DNA repair foci, so FAND2 foci restoration constitutes therefore a hallmark in ICL repair. Moreover, the reversion of MMC hyper-sensitivity in FA-A cells (Figure 67) confirms the correction of the FA phenotype.

One possibility for enhancing the efficacy of gene targeting is based on the use of small molecules capable of enhancing the HR in the cell. Several groups have studied this possibility using small molecules such as L755507, which has been reported to increase the HDR efficiency in different cells types treated with CRISPR/Cas9 nucleases[147]. Moreover other groups has inhibited key components of NHEJ such as DNA ligase IV, during their gene targeting experiments in order to increase the efficiency of HDR[148,149]. In our experiments, the repair of nuclease derived-DSBs by NHEJ was, as expected, higher than that mediated by HDR in HSPCs (Table 12). HSPCs are more quiescent cells[318], and the HR machinery is only turned on during the S-phase of the cell cycle. As it has been already mentioned, one possible approach to increase the efficiency of HDR is to use nickase variants instead of nucleases[308,309]. This approach was addressed in FA cells in the Osborn and coworkers' study, in which they use a CRISPR/Cas9 nickase to promote HDR in FA-C cells[292]. However, in this study although the nickase reduces the percentage of NHEJ events to repair the DSBs, it also decreases the efficiency of HDR with respect to the nuclease.

In 2012 Moldovan and coworkers published that the inhibition of the antirecombinase PARI in FANCD1/BRCA2 deficient fibroblasts increased the efficiency of HR[150]. In healthy cells, during S-phase of cell cycle, PARI interacts with PCNA to release RAD51 from ssDNA, avoiding recombination events at replication forks. This prevents hyper-recombination events that lead to genomic instability. However, in FANCD1 deficient cells, where HR is impaired, PARI

inhibition leads to an increase in RAD51 levels and consequently an improvement in HR efficiency that finally leads to an increase in genomic stability. In our experiments, we demonstrated that siRNA inhibited PARI in both FA-A LCLs (68.12±0.08% of inhibition) and also in FA-A fibroblasts (68.51±0.12%). We also showed an increase in RAD51 foci formation, either in the presence or absence of DNA damage, in PARI interfered FA-A cells (Figure 69 a), suggesting that HR is increased by PARI inhibition not only in FA-D1 cells[150], but also in FA-A cells. It is also important to consider that the inhibition of this gene in FA-A cells decreased the intrinsic genomic instability of these cells (Figure 69 b, c) confirming the protective effects of PARI inhibition in FA-A cells, as was also shown by Moldovan et al[150] in FA-D1 cells. However, when gene targeting was tested in PARI interfered FA-A cells (but not in HD cells), only mild, non-significant, increases in gene editing efficiency were observed (Figure 70 b, Figure 72 b), probably due to the variability observed between experiments. Although we cannot demonstrate robustly that PARI inhibition increases gene targeting in FA-A cells, the use of PARI-inhibiting molecules may constitute less toxic a more reproducible approach to demonstrate the role of PARI inhibition in gene targeting in FA-A cells, and also could be translated easily to HSPCs.

In our final experiments, we moved from LCLs to hematopoietic progenitor cells from FA-A patients. In these studies, we used a very small proportion of mPB $CD34^+$ cells that remained in cell collection bags and tubes from the CliniMACS® system used to purify $CD34^+$ cells from FA-A patients included in the FANCOSTEM mobilization trial. This procedure was performed after the informed consent of patients or their parents and with the approval of the ethic committee of the hospitals. Here, we have demonstrated for the first time that gene targeting is feasible in FA-A $CD34^+$ cells from four different patients (Figure 73) using a safe harbor strategy. Additionally, we demonstrate this approach corrects the phenotype of these cells, as we were able to partially restore the MMC sensitivity of these cells after gene targeting with the therapeutic donor and also we confirmed the specific integration in the *AAVS1* locus by PCR after MMC selection (Figure 74). Gene targeting efficiency has been also proven in $CD34^+$ cells from patients with other hematological disorders such as SCID-X1, in which gene targeting was demonstrated in bone-marrow derived $CD34^+$ cells from one patient[113] and sickle cell disease, in which gene editing was also shown in BM $CD34^+$ cell aspirates[115]. Ongoing experiments aim to demonstrate the feasibility of gene edited FA-A $CD34^+$ cells to engraft in immunodeficient mice.

Taking into account all our data, we propose that FA-A hematopoietic cells, can be corrected by gene editing. Our proposed strategy of gene targeting in a safe harbor locus should abrogate the eventual risk of insertional mutagenesis. Moreover, the proliferative advantage that FA corrected cells present over non-corrected ones[250] suggest that gene editing efficiencies obtained in our study might ameliorate the BMF of FA patients. The somatic mosaicism observed in FA patients, in whom an *in vivo* spontaneous reversion of a single HSPC corrects the hematopoietic phenotype of the patients support this idea[5]. However, considering the low number of $CD34^+$ cells usually obtained from FA patients, the improvement in gene targeting efficiencies will be essential to assure the transplant of a relatively high number of corrected cells in the patients and their future successful in the clinics.

CONCLUSIONS

1.

Aiming at the development of new targeted gene therapy approaches in FA-A cells, we have constructed an IDLV harboring the *FANCA* gene under the regulation of the PGK promoter and flanked by the homology arms corresponding to the first intron of the *AAVS1* locus.

2.

The combined use of zinc finger nucleases specific for the *AAVS1* locus together with IDLV donors, either harboring the *GFP* reporter or the *FANCA* therapeutic genes, facilitated the targeted integration of these genes into LCLs and cord blood hematopoietic stem cells from healthy donors.

3.

The transplantation of targeted $CD34^+$ cells into NSG-SGM3 immunodeficient mice allowed us to demonstrate the specific integration of the *GFP* reporter gene in the *AAVS1* safe harbor locus in human long term hematopoietic repopulating cells at average efficacies of 11%. This value constitutes the highest level of gene targeting reported so far in human hematopoietic stem cells.

4.

The use of gold nanoparticles improved the transduction efficiency of lentiviral vectors in hematopoietic progenitors and short-term repopulating cells. However, the combination of gold nanoparticles with nucleofection was very toxic for $CD34^+$ cells, preventing their application in our gene targeting approaches.

5.

The gene targeting efficiency observed in FA-A LCLs was only modestly reduced as compared with that observed in either their corrected counterparts or in healthy donor LCLs. This demonstrates for the first time that FANCA protein, a member of the core complex of the FA/BRCA pathway, is not essential to conduct gene targeting in the genome of hematopoietic cells.

6.

The targeted integration of *FANCA* in the *AAVS1* locus of FA-A LCLs restores the formation of FANCD2 foci and the correction of the hyper-sensitivity of these cells to mitomycin C.

7.

The inhibition of the antirecombinase PARI increased the proportion of FA-A cells with RAD51 foci, but did not significantly improve the efficacy of gene targeting, either in LCLs or in fibroblasts from FA-A patients.

8.

Our gene targeting approach showed for the first time that gene editing is feasible in hematopoietic stem and progenitor cells from FA-A patients, leading to the partial reversion of their characteristic hyper-sensitivity to mitomycin C.

Altogether the results presented in this study demonstrate for the first time the feasibility of conducting targeted gene therapy in hematopoietic progenitor cells from FA patients characterized by DNA repair defects. These results open new perspectives for the development of the safest gene therapy approach that nowadays can be envisaged for the treatment of monogenic diseases. Our data also reflects that further improvements in gene editing efficacy should be developed before these strategies could be applied for the treatment of patients affected by FA or similar devastating diseases.

1.

Con el objetivo de desarrollar nuevas aproximaciones de terapia génica de edición en células AF-A, hemos construido un vector lentiviral defectivo en integración (VLDI) que porta el gen *FANCA* regulado por el promotor PGK y flanqueado por los brazos de homología correspondientes al primer intrón del locus *AAVS1*.

2.

La combinación de nucleasas de dedos de zinc específicas para el locus *AAVS1* con VLDIs que portan el gen marcador *GFP* o el gen terapéutico *FANCA*, facilita la integración dirigida de estos genes en células linfoblastoides (LCLs), así como en células $CD34^+$ de sangre de cordón umbilical de donantes sanos.

3.

El trasplante de células $CD34^+$ editadas genéticamente en ratones inmunodeficientes NSG-SGM3, nos permitió demostrar la integración específica del gen marcador *GFP* en el locus seguro *AAVS1* en el 11% de las células hematopoyéticas con capacidad de reconstitución a largo plazo. Este es el nivel más alto de edición génica obtenido hasta el momento en células madre hematopoyéticas humanas.

4.

El uso de nanopartículas de oro mejoró la eficacia de transducción de los vectores lentivirales en progenitores hematopoyéticos y en células con capacidad de repoblación a corto plazo. Sin embargo, la combinación de nanopartículas de oro con nucleofección resultó muy tóxica para las células $CD34^+$, lo que impidió su aplicación en nuestra aproximación de edición génica.

5.

La eficiencia de edición génica en LCLs de AF-A es sólo ligeramente inferior a la observada en las mismas células corregidas genéticamente o en LCLs de donantes sanos. Esto demuestra por primera vez que la proteína FANCA, un miembro del complejo *core* de la ruta AF/BRCA, no es esencial para llevar a cabo terapia génica de edición mediada por recombinación homóloga en células hematopoyéticas.

6.

La integración dirigida del gen *FANCA* en el locus *AAVS1* en LCLs AF-A recuperó la formación de focos de reparación de FANCD2, así como la corrección de su hipersensibilidad a mitomicina C.

7.

La inhibición de la proteína anti-recombinasa PARI aumentó la proporción de células AF-A con focos de reparación de RAD51, pero no aumentó la eficacia de edición génica en LCLs ni en fibroblastos de pacientes con AF-A.

8.

Nuestra aproximación experimental demuestra por primera vez que la edición génica es factible en células progenitoras hematopoyéticas obtenidas de pacientes con AF-A, conduciendo a la reversión parcial de su característica hipersensibilidad a mitomicina C.

En conjunto, los resultados presentados en este estudio demuestran por primera vez la posibilidad de realizar terapia génica de edición en células progenitoras hematopoyéticas de pacientes con AF, caracterizados por defectos en la reparación del ADN. Estos resultados abren nuevas perspectivas para el desarrollo de aproximaciones de terapia génica más segura para el tratamiento de diferentes enfermedades monogénicas. Nuestros datos también reflejan que para la aplicación clínica de estas estrategias en pacientes con AF y otras enfermedades similares, será necesario continuar mejorando su eficacia en las células diana afectadas.

BIBLIOGRAPHY

1 Ghosh, S., Thrasher, A. J. & Gaspar, H. B. Gene therapy for monogenic disorders of the bone marrow. *British journal of haematology*, doi:10.1111/bjh.13520 (2015).

2 Naldini, L. Ex vivo gene transfer and correction for cell-based therapies. *Nature reviews. Genetics* **12**, 301-315, doi:10.1038/nrg2985 (2011).

3 Kaufmann, K. B., Buning, H., Galy, A., Schambach, A. & Grez, M. Gene therapy on the move. *EMBO molecular medicine* **5**, 1642-1661, doi:10.1002/emmm.201202287 (2013).

4 Kay, M. A. State-of-the-art gene-based therapies: the road ahead. *Nature reviews. Genetics* **12**, 316-328, doi:10.1038/nrg2971 (2011).

5 Gross, M. *et al.* Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenetic and genome research* **98**, 126-135, doi:69805 (2002).

6 Aiuti, A. *et al.* Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *The New England journal of medicine* **360**, 447-458, doi:10.1056/NEJMoa0805817 (2009).

7 Gaspar, H. B. *et al.* Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. *Science translational medicine* **3**, 97ra80, doi:10.1126/scitranslmed.3002716 (2011).

8 Candotti, F. *et al.* Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans. *Blood* **120**, 3635-3646, doi:10.1182/blood-2012-02-400937 (2012).

9 Hacein-Bey-Abina, S. *et al.* Efficacy of gene therapy for X-linked severe combined immunodeficiency. *The New England journal of medicine* **363**, 355-364, doi:10.1056/NEJMoa1000164 (2010).

10 Fischer, A., Hacein-Bey-Abina, S. & Cavazzana-Calvo, M. 20 years of gene therapy for SCID. *Nature immunology* **11**, 457-460, doi:10.1038/ni0610-457 (2010).

11 Gaspar, H. B. *et al.* Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Science translational medicine* **3**, 97ra79, doi:10.1126/scitranslmed.3002715 (2011).

12 Mukherjee, S. & Thrasher, A. J. Gene therapy for PIDs: progress, pitfalls and prospects. *Gene* **525**, 174-181, doi:10.1016/j.gene.2013.03.098 (2013).

13 Hacein-Bey-Abina, S. *et al.* A modified gamma-retrovirus vector for X-linked severe combined immunodeficiency. *The New England journal of medicine* **371**, 1407-1417, doi:10.1056/NEJMoa1404588 (2014).

14 Touzot, F. *et al.* Faster T-cell development following gene therapy compared with haploidentical HSCT in the treatment of SCID-X1. *Blood* **125**, 3563-3569, doi:10.1182/blood-2014-12-616003 (2015).

15 Grez, M. *et al.* Gene therapy of chronic granulomatous disease: the engraftment dilemma. *Molecular therapy : the journal of the American Society of Gene Therapy* **19**, 28-35, doi:10.1038/mt.2010.232 (2011).

16 Boztug, K. *et al.* Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *The New England journal of medicine* **363**, 1918-1927, doi:10.1056/NEJMoa1003548 (2010).

17 Braun, C. J. *et al.* Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Science translational medicine* **6**, 227ra233, doi:10.1126/scitranslmed.3007280 (2014).

18 Aiuti, A. *et al.* Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* **341**, 1233151, doi:10.1126/science.1233151 (2013).

19 Castiello, M. C. *et al.* B-cell reconstitution after lentiviral vector-mediated gene therapy in patients with Wiskott-Aldrich syndrome. *The Journal of allergy and clinical immunology* **136**, 692-702 e692, doi:10.1016/j.jaci.2015.01.035 (2015).

20 Hacein-Bey Abina, S. *et al.* Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. *Jama* **313**, 1550-1563, doi:10.1001/jama.2015.3253 (2015).

21 Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**, 318-322, doi:10.1038/nature09328 (2010).

22 Cartier, N. *et al.* Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326**, 818-823, doi:10.1126/science.1171242 (2009).

23 Biffi, A. *et al.* Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **341**, 1233158, doi:10.1126/science.1233158 (2013).

24 Gabriel, R., Schmidt, M. & von Kalle, C. Integration of retroviral vectors. *Current opinion in immunology* **24**, 592-597, doi:10.1016/j.coi.2012.08.006 (2012).

25 Cavazza, A., Moiani, A. & Mavilio, F. Mechanisms of retroviral integration and mutagenesis. *Human gene therapy* **24**, 119-131, doi:10.1089/hum.2012.203 (2013).

26 Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419, doi:10.1126/science.1088547 (2003).

27 Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of clinical investigation* **118**, 3143-3150, doi:10.1172/JCI35798 (2008).

28 De Palma, M. *et al.* Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. *Blood* **105**, 2307-2315, doi:10.1182/blood-2004-03-0798 (2005).

29 Cesana, D. *et al.* Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *The Journal of clinical investigation* **122**, 1667-1676, doi:10.1172/JCI62189 (2012).

30 Moiani, A. *et al.* Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *The Journal of clinical investigation* **122**, 1653-1666, doi:10.1172/JCI61852 (2012).

31 Cathomen, T. & Joung, J. K. Zinc-finger nucleases: the next generation emerges. *Molecular therapy : the journal of the American Society of Gene Therapy* **16**, 1200-1207, doi:10.1038/mt.2008.114 (2008).

32 Maggio, I. & Goncalves, M. A. Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends Biotechnol* **33**, 280-291, doi:10.1016/j.tibtech.2015.02.011 (2015).

33 Delacote, F. & Lopez, B. S. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model. *Cell cycle* **7**, 33-38 (2008).

34 Agarwal, S., Tafel, A. A. & Kanaar, R. DNA double-strand break repair and chromosome translocations. *DNA repair* **5**, 1075-1081, doi:10.1016/j.dnarep.2006.05.029 (2006).

35 Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annual review of genetics* **40**, 363-383, doi:10.1146/annurev.genet.40.110405.090451 (2006).

36 Fattah, F. J., Lichter, N. F., Fattah, K. R., Oh, S. & Hendrickson, E. A. Ku70, an essential gene, modulates the frequency of rAAV-mediated gene targeting in human somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8703-8708, doi:10.1073/pnas.0712060105 (2008).

37 Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18**, 134-147, doi:10.1038/cr.2007.111 (2008).

38 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

39 Meek, K., Gupta, S., Ramsden, D. A. & Lees-Miller, S. P. The DNA-dependent protein kinase: the director at the end. *Immunological reviews* **200**, 132-141, doi:10.1111/j.0105-2896.2004.00162.x (2004).

40 Weterings, E. & Chen, D. J. The endless tale of non-homologous end-joining. *Cell Res* **18**, 114-124, doi:10.1038/cr.2008.3 (2008).

41 Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Translational cancer research* **2**, 130-143, doi:10.3978/j.issn.2218-676X.2013.04.02 (2013).

42 Iliakis, G. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiother Oncol* **92**, 310-315, doi:10.1016/j.radonc.2009.06.024 (2009).

43 Mladenov, E. & Iliakis, G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutation research* **711**, 61-72, doi:10.1016/j.mrfmmm.2011.02.005 (2011).

44 Wang, M. *et al.* PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research* **34**, 6170-6182, doi:10.1093/nar/gkl840 (2006).

45 Zha, S., Boboila, C. & Alt, F. W. Mre11: roles in DNA repair beyond homologous recombination. *Nature structural & molecular biology* **16**, 798-800, doi:10.1038/nsmb0809-798 (2009).

46 Lee-Theilen, M., Matthews, A. J., Kelly, D., Zheng, S. & Chaudhuri, J. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nature structural & molecular biology* **18**, 75-79, doi:10.1038/nsmb.1942 (2011).

47 Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA ligases: structural and functional insights. *Annual review of biochemistry* **77**, 313-338, doi:10.1146/annurev.biochem.77.061306.123941 (2008).

48 Sallmyr, A., Tomkinson, A. E. & Rassool, F. V. Up-regulation of WRN and DNA ligase IIIalpha in chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks. *Blood* **112**, 1413-1423, doi:10.1182/blood-2007-07-104257 (2008).

49 Mladenov, E., Magin, S., Soni, A. & Iliakis, G. DNA double-strand break repair as determinant of cellular radiosensitivity to killing and target in radiation therapy. *Front Oncol* **3**, 113, doi:10.3389/fonc.2013.00113 (2013).

50 Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews. Genetics* **2**, 292-301, doi:10.1038/35066075 (2001).

51 Aylon, Y. & Kupiec, M. New insights into the mechanism of homologous recombination in yeast. *Mutation research* **566**, 231-248, doi:10.1016/j.mrrev.2003.10.001 (2004).

52 Bernstein, K. A. & Rothstein, R. At loose ends: resecting a double-strand break. *Cell* **137**, 807-810, doi:10.1016/j.cell.2009.05.007 (2009).

53 Aly, A. & Ganesan, S. BRCA1, PARP, and 53BP1: conditional synthetic lethality and synthetic viability. *J Mol Cell Biol* **3**, 66-74, doi:10.1093/jmcb/mjq055 (2011).

54 Escribano-Diaz, C. *et al.* A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Molecular cell* **49**, 872-883, doi:10.1016/j.molcel.2013.01.001 (2013).

55 Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7720-7725, doi:10.1073/pnas.1213431110 (2013).

56 Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509-514, doi:10.1038/nature06337 (2007).

57 Grabarz, A. *et al.* A role for BLM in double-strand break repair pathway choice: prevention of CtIP/Mre11-mediated alternative nonhomologous end-joining. *Cell reports* **5**, 21-28, doi:10.1016/j.celrep.2013.08.034 (2013).

58 Garate, Z., Davis, B. R., Quintana-Bustamante, O. & Segovia, J. C. New frontier in regenerative medicine: site-specific gene correction in patient-specific induced pluripotent stem cells. *Human gene therapy* **24**, 571-583, doi:10.1089/hum.2012.251 (2013).

59 Castella, M. *et al.* Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood* **117**, 3759-3769, doi:10.1182/blood-2010-08-299917 (2011).

60 Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nature reviews. Cancer* **12**, 51-58, doi:10.1038/nrc3179 (2012).

61 Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).

62 Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New England journal of medicine* **370**, 901-910, doi:10.1056/NEJMoa1300662 (2014).

63 Rottman, J. B. *et al.* Cellular localization of the chemokine receptor CCR5. Correlation to cellular targets of HIV-1 infection. *The American journal of pathology* **151**, 1341-1351 (1997).

64 Lim, J. K., Glass, W. G., McDermott, D. H. & Murphy, P. M. CCR5: no longer a "good for nothing" gene--chemokine control of West Nile virus infection. *Trends in immunology* **27**, 308-312, doi:10.1016/j.it.2006.05.007 (2006).

65 Lombardo, A. *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nature methods* **8**, 861-869, doi:10.1038/nmeth.1674 (2011).

66 Mussolino, C. *et al.* TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research* **42**, 6762-6773, doi:10.1093/nar/gku305 (2014).

67 Kotin, R. M. *et al.* Site-specific integration by adeno-associated virus. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2211-2215 (1990).

68 Henckaerts, E. & Linden, R. M. Adeno-associated virus: a key to the human genome? *Future virology* **5**, 555-574, doi:10.2217/fvl.10.48 (2010).

69 Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature biotechnology* **27**, 851-857, doi:10.1038/nbt.1562 (2009).

70 DeKelver, R. C. *et al.* Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome research* **20**, 1133-1142, doi:10.1101/gr.106773.110 (2010).

71 Zou, J. *et al.* Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood* **117**, 5561-5572, doi:10.1182/blood-2010-12-328161 (2011).

72 Ramachandra, C. J. *et al.* Efficient recombinase-mediated cassette exchange at the AAVS1 locus in human embryonic stem cells using baculoviral vectors. *Nucleic acids research* **39**, e107, doi:10.1093/nar/gkr409 (2011).

73 Smith, J. R. *et al.* Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem cells* **26**, 496-504, doi:10.1634/stemcells.2007-0039 (2008).

74 Yang, L. *et al.* Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* **453**, 524-528, doi:10.1038/nature06894 (2008).

75 Rio, P. *et al.* Targeted gene therapy and cell reprogramming in Fanconi anemia. *EMBO molecular medicine* **6**, 835-848, doi:10.15252/emmm.201303374 (2014).

76 Ogata, T., Kozuka, T. & Kanda, T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *Journal of virology* **77**, 9000-9007 (2003).

77 Elliott, B. & Jasin, M. Double-strand breaks and translocations in cancer. *Cell Mol Life Sci* **59**, 373-385 (2002).

78 Grizot, S. *et al.* Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic acids research* **38**, 2006-2018, doi:10.1093/nar/gkp1171 (2010).

79 Paques, F. & Duchateau, P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Current gene therapy* **7**, 49-66 (2007).

80 Chevalier, B. S. & Stoddard, B. L. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic acids research* **29**, 3757-3774 (2001).

81 Epinat, J. C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic acids research* **31**, 2952-2962 (2003).

82 Riviere, J. *et al.* Variable correction of Artemis deficiency by I-Sce1-meganuclease-assisted homologous recombination in murine hematopoietic stem cells. *Gene therapy* **21**, 529-532, doi:10.1038/gt.2014.20 (2014).

83 Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436, doi:10.1146/annurev-phyto-080508-081936 (2010).

84 Mussolino, C. & Cathomen, T. TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol* **23**, 644-650, doi:10.1016/j.copbio.2012.01.013 (2012).

85 Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-723, doi:10.1126/science.1215670 (2012).

86 Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).

87 Bitinaite, J., Wah, D. A., Aggarwal, A. K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10570-10575 (1998).

88 Smith, J. *et al.* Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic acids research* **28**, 3361-3369 (2000).

89 Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research* **39**, 9283-9293, doi:10.1093/nar/gkr597 (2011).

90 Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* **45**, 273-297, doi:10.1146/annurev-genet-110410-132430 (2011).

91 Al-Attar, S., Westra, E. R., van der Oost, J. & Brouns, S. J. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol Chem* **392**, 277-289, doi:10.1515/BC.2011.042 (2011).

92 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

93 Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31**, 230-232, doi:10.1038/nbt.2507 (2013).

94 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).

95 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

96 Maeder, M. L. *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* **31**, 294-301, doi:10.1016/j.molcel.2008.06.016 (2008).

97 Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F. & Joung, J. K. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc* **4**, 1471-1501, doi:10.1038/nprot.2009.98 (2009).

98 Pabo, C. O., Peisach, E. & Grant, R. A. Design and selection of novel Cys2His2 zinc finger proteins. *Annual review of biochemistry* **70**, 313-340, doi:10.1146/annurev.biochem.70.1.313 (2001).

99 Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology* **29**, 816-823, doi:10.1038/nbt.1948 (2011).

100 Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods* **8**, 765-770, doi:10.1038/nmeth.1670 (2011).

101 Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature biotechnology* **25**, 778-785, doi:10.1038/nbt1319 (2007).

102 Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature methods* **8**, 74-79, doi:10.1038/nmeth.1539 (2011).

103 Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156-1160 (1996).

104 Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology* **26**, 695-701, doi:10.1038/nbt1398 (2008).

105 Carbery, I. D. *et al.* Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186**, 451-459, doi:10.1534/genetics.110.117002 (2010).

106 Geurts, A. M. *et al.* Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433, doi:10.1126/science.1172447 (2009).
107 Mashimo, T. *et al.* Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PloS one* **5**, e8870, doi:10.1371/journal.pone.0008870 (2010).
108 Yang, D. *et al.* Production of apolipoprotein C-III knockout rabbits using zinc finger nucleases. *J Vis Exp*, e50957, doi:10.3791/50957 (2013).
109 Hauschild, J. *et al.* Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12013-12017, doi:10.1073/pnas.1106422108 (2011).
110 Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature biotechnology* **26**, 808-816, doi:10.1038/nbt1410 (2008).
111 Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature biotechnology* **28**, 839-847, doi:10.1038/nbt.1663 (2010).
112 Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature biotechnology* **25**, 1298-1306, doi:10.1038/nbt1353 (2007).
113 Genovese, P. *et al.* Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235-240, doi:10.1038/nature13420 (2014).
114 Merling, R. K. *et al.* An AAVS1-targeted minigene platform for correction of iPSCs from all five types of chronic granulomatous disease. *Molecular therapy : the journal of the American Society of Gene Therapy* **23**, 147-157, doi:10.1038/mt.2014.195 (2015).
115 Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125**, 2597-2604, doi:10.1182/blood-2014-12-615948 (2015).
116 Skipper, K. A. & Mikkelsen, J. G. Delivering the goods for genome engineering and editing. *Human gene therapy*, doi:10.1089/hum.2015.063 (2015).
117 Papapetrou, E. P., Zoumbos, N. C. & Athanassiadou, A. Genetic modification of hematopoietic stem cells with nonviral systems: past progress and future prospects. *Gene therapy* **12 Suppl 1**, S118-130, doi:10.1038/sj.gt.3302626 (2005).
118 Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-651, doi:10.1038/nature03556 (2005).
119 Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763, doi:10.1126/science.1078395 (2003).
120 Orlando, S. J. *et al.* Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic acids research* **38**, e152, doi:10.1093/nar/gkq512 (2010).
121 Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature methods* **8**, 753-755, doi:10.1038/nmeth.1653 (2011).
122 Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3055-3060, doi:10.1073/pnas.0611478104 (2007).
123 Kuhn, A. N. *et al.* mRNA as a versatile tool for exogenous protein expression. *Current gene therapy* **12**, 347-361 (2012).
124 Sahin, U., Kariko, K. & Tureci, O. mRNA-based therapeutics--developing a new class of drugs. *Nat Rev Drug Discov* **13**, 759-780, doi:10.1038/nrd4278 (2014).
125 Shi, N. Q., Qi, X. R., Xiang, B. & Zhang, Y. A survey on "Trojan Horse" peptides: opportunities, issues and controlled entry to "Troy". *J Control Release* **194**, 53-70, doi:10.1016/j.jconrel.2014.08.014 (2014).
126 Wadia, J. S., Stan, R. V. & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nature medicine* **10**, 310-315, doi:10.1038/nm996 (2004).

127 Gaj, T., Guo, J., Kato, Y., Sirk, S. J. & Barbas, C. F., 3rd. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature methods* **9**, 805-807, doi:10.1038/nmeth.2030 (2012).

128 Chen, Z. *et al.* Receptor-mediated delivery of engineered nucleases for genome modification. *Nucleic acids research* **41**, e182, doi:10.1093/nar/gkt710 (2013).

129 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **24**, 1012-1019, doi:10.1101/gr.171322.113 (2014).

130 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).

131 Cai, Y., Bak, R. O. & Mikkelsen, J. G. Targeted genome editing by lentiviral protein transduction of zinc-finger and TAL-effector nucleases. *Elife* **3**, e01911, doi:10.7554/eLife.01911 (2014).

132 Mock, U. *et al.* Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. *Scientific reports* **4**, 6409, doi:10.1038/srep06409 (2014).

133 Lei, Y. *et al.* Gene editing of human embryonic stem cells via an engineered baculoviral vector carrying zinc-finger nucleases. *Molecular therapy : the journal of the American Society of Gene Therapy* **19**, 942-950, doi:10.1038/mt.2011.12 (2011).

134 Anguela, X. M. *et al.* Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood* **122**, 3283-3287, doi:10.1182/blood-2013-04-497354 (2013).

135 Handel, E. M. *et al.* Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors. *Human gene therapy* **23**, 321-329, doi:10.1089/hum.2011.140 (2012).

136 Ellis, B. L., Hirsch, M. L., Porter, S. N., Samulski, R. J. & Porteus, M. H. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food and Drug Administration-approved drugs. *Gene therapy* **20**, 35-42, doi:10.1038/gt.2011.211 (2013).

137 Rahman, S. H. *et al.* The nontoxic cell cycle modulator indirubin augments transduction of adeno-associated viral vectors and zinc-finger nuclease-mediated gene targeting. *Human gene therapy* **24**, 67-77, doi:10.1089/hum.2012.168 (2013).

138 Mazurier, F., Gan, O. I., McKenzie, J. L., Doedens, M. & Dick, J. E. Lentivector-mediated clonal tracking reveals intrinsic heterogeneity in the human hematopoietic stem cell compartment and culture-induced stem cell impairment. *Blood* **103**, 545-552, doi:10.1182/blood-2003-05-1558 (2004).

139 Wang, C. X. *et al.* Rapamycin relieves lentiviral vector transduction resistance in human and mouse hematopoietic stem cells. *Blood* **124**, 913-923, doi:10.1182/blood-2013-12-546218 (2014).

140 Petrillo, C. *et al.* Cyclosporin a and rapamycin relieve distinct lentiviral restriction blocks in hematopoietic stem and progenitor cells. *Molecular therapy : the journal of the American Society of Gene Therapy* **23**, 352-362, doi:10.1038/mt.2014.193 (2015).

141 Conde, J. *et al.* Revisiting 30 years of biofunctionalization and surface chemistry of inorganic nanoparticles for nanomedicine. *Front Chem* **2**, 48, doi:10.3389/fchem.2014.00048 (2014).

142 Bao, C. *et al.* A promising road with challenges: where are gold nanoparticles in translational research? *Nanomedicine (Lond)* **9**, 2353-2370, doi:10.2217/nnm.14.155 (2014).

143 Lee, S. K., Han, M. S., Asokan, S. & Tung, C. H. Effective gene silencing by multilayered siRNA-coated gold nanoparticles. *Small* **7**, 364-370, doi:10.1002/smll.201001314 (2011).

144 Conde, J. *et al.* In vivo tumor targeting via nanoparticle-mediated therapeutic siRNA coupled to inflammatory response in lung cancer mouse models. *Biomaterials* **34**, 7744-7753, doi:10.1016/j.biomaterials.2013.06.041 (2013).

145 Conde, J. *et al.* Design of multifunctional gold nanoparticles for in vitro and in vivo gene silencing. *ACS nano* **6**, 8316-8324, doi:10.1021/nn3030223 (2012).

146 Kamei, K. *et al.* Direct cell entry of gold/iron-oxide magnetic nanoparticles in adenovirus mediated gene delivery. *Biomaterials* **30**, 1809-1814, doi:10.1016/j.biomaterials.2008.12.015 (2009).
147 Yu, C. *et al.* Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell stem cell* **16**, 142-147, doi:10.1016/j.stem.2015.01.003 (2015).
148 Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548, doi:10.1038/nbt.3198 (2015).
149 Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-542, doi:10.1038/nbt.3190 (2015).
150 Moldovan, G. L. *et al.* Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Molecular cell* **45**, 75-86, doi:10.1016/j.molcel.2011.11.010 (2012).
151 Kee, Y. & D'Andrea, A. D. Molecular pathogenesis and clinical management of Fanconi anemia. *The Journal of clinical investigation* **122**, 3799-3806, doi:10.1172/JCI58321 (2012).
152 Fanconi, G. Familiäre infantile perniziosaartige Anämie (perniciziöses Blutbild und Konstitution). *Jahrbuch für Kinderheikunde und Erziehung (Wien)* **117**, 257-280 (1927).
153 Lobitz, S. & Velleuer, E. Guido Fanconi (1892-1979): a jack of all trades. *Nature reviews. Cancer* **6**, 893-898, doi:10.1038/nrc2009 (2006).
154 Bogliolo, M. & Surralles, J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr Opin Genet Dev* **33**, 32-40, doi:10.1016/j.gde.2015.07.002 (2015).
155 Navarro, S., Rio, P. & Bueren, J. A. Perspectives on Gene Therapy for Fanconi Anemia. *Expert Opin Orphan Drugs* **3**, 899-910, doi:10.1517/21678707.2015.1062752 (2015).
156 Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nature reviews. Cancer* **11**, 467-480, doi:10.1038/nrc3088 (2011).
157 Kim, H. & D'Andrea, A. D. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes & development* **26**, 1393-1408, doi:10.1101/gad.195248.112 (2012).
158 Kottemann, M. C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**, 356-363, doi:10.1038/nature11863 (2013).
159 Clauson, C., Scharer, O. D. & Niedernhofer, L. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harbor perspectives in medicine* **3**, a012732 (2013).
160 Garaycoechea, J. I. & Patel, K. J. Why does the bone marrow fail in Fanconi anemia? *Blood* **123**, 26-34, doi:10.1182/blood-2013-09-427740 (2014).
161 Auerbach, A. D. Fanconi anemia and its diagnosis. *Mutation research* **668**, 4-10, doi:10.1016/j.mrfmmm.2009.01.013 (2009).
162 Singh, T. R. *et al.* Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* **114**, 174-180, doi:10.1182/blood-2009-02-207811 (2009).
163 Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS genetics* **10**, e1004494, doi:10.1371/journal.pgen.1004494 (2014).
164 Fanconi anaemia/Breast cancer, c. Positional cloning of the Fanconi anaemia group A gene. *Nature genetics* **14**, 324-328, doi:10.1038/ng1196-324 (1996).
165 Foe, J. R. *et al.* Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nature genetics* **14**, 488, doi:10.1038/ng1296-488 (1996).
166 Meetei, A. R. *et al.* X-linked inheritance of Fanconi anemia complementation group B. *Nature genetics* **36**, 1219-1224, doi:10.1038/ng1458 (2004).
167 Strathdee, C. A., Gavish, H., Shannon, W. R. & Buchwald, M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **358**, 434, doi:10.1038/358434a0 (1992).
168 Howlett, N. G. *et al.* Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**, 606-609, doi:10.1126/science.1073834 (2002).

169 Tavtigian, S. V. *et al.* The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nature genetics* **12**, 333-337, doi:10.1038/ng0396-333 (1996).
170 Timmers, C. *et al.* Positional cloning of a novel Fanconi anemia gene, FANCD2. *Molecular cell* **7**, 241-248 (2001).
171 de Winter, J. P. *et al.* Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *American journal of human genetics* **67**, 1306-1308, doi:10.1016/S0002-9297(07)62959-0 (2000).
172 de Winter, J. P. *et al.* The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM. *Nature genetics* **24**, 15-16, doi:10.1038/71626 (2000).
173 de Winter, J. P. *et al.* The Fanconi anaemia group G gene FANCG is identical with XRCC9. *Nature genetics* **20**, 281-283, doi:10.1038/3093 (1998).
174 Garcia-Higuera, I., Kuang, Y., Naf, D., Wasik, J. & D'Andrea, A. D. Fanconi anemia proteins FANCA, FANCC, and FANCG/XRCC9 interact in a functional nuclear complex. *Molecular and cellular biology* **19**, 4866-4873 (1999).
175 Levitus, M. *et al.* Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* **103**, 2498-2503, doi:10.1182/blood-2003-08-2915 (2004).
176 Dorsman, J. C. *et al.* Identification of the Fanconi anemia complementation group I gene, FANCI. *Cellular oncology : the official journal of the International Society for Cellular Oncology* **29**, 211-218 (2007).
177 Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nature structural & molecular biology* **14**, 564-567, doi:10.1038/nsmb1252 (2007).
178 Smogorzewska, A. *et al.* Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289-301, doi:10.1016/j.cell.2007.03.009 (2007).
179 Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nature structural & molecular biology* **15**, 1138-1146, doi:10.1038/nsmb.1504 (2008).
180 Meetei, A. R. *et al.* A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature genetics* **35**, 165-170, doi:10.1038/ng1241 (2003).
181 Meetei, A. R. *et al.* A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nature genetics* **37**, 958-963, doi:10.1038/ng1626 (2005).
182 Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature genetics* **39**, 165-167, doi:10.1038/ng1959 (2007).
183 Reid, S. *et al.* Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nature genetics* **39**, 162-164, doi:10.1038/ng1947 (2007).
184 Xia, B. *et al.* Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nature genetics* **39**, 159-161, doi:10.1038/ng1942 (2007).
185 Zhang, F. *et al.* PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Current biology : CB* **19**, 524-529, doi:10.1016/j.cub.2009.02.018 (2009).
186 Vaz, F. *et al.* Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nature genetics* **42**, 406-409, doi:10.1038/ng.570 (2010).
187 Stoepker, C. *et al.* SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nature genetics* **43**, 138-141, doi:10.1038/ng.751 (2011).
188 Kim, Y. *et al.* Mutations of the SLX4 gene in Fanconi anemia. *Nature genetics* **43**, 142-146, doi:10.1038/ng.750 (2011).
189 Bogliolo, M. *et al.* Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *American journal of human genetics* **92**, 800-806, doi:10.1016/j.ajhg.2013.04.002 (2013).
190 Wang, A. T. *et al.* A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Molecular cell* **59**, 478-490, doi:10.1016/j.molcel.2015.07.009 (2015).
191 Sawyer, S. L. *et al.* Biallelic Mutations in BRCA1 Cause a New Fanconi Anemia Subtype. *Cancer discovery*, doi:10.1158/2159-8290.CD-14-1156 (2014).

192 Hira, A. *et al.* Mutations in the Gene Encoding the E2 Conjugating Enzyme UBE2T Cause Fanconi Anemia. *American journal of human genetics* **96**, 1001-1007, doi:10.1016/j.ajhg.2015.04.022 (2015).

193 Virts, E. L. *et al.* AluY-mediated Germline Deletion, Duplication and Somatic Stem Cell Reversion in UBE2T Defines a New Subtype of Fanconi Anemia. *Human molecular genetics*, doi:10.1093/hmg/ddv227 (2015).

194 Alter, B. P. & Kupfer, G. in *GeneReviews(R)* (eds R. A. Pagon *et al.*) (1993).

195 Shen, X. *et al.* Recruitment of fanconi anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. *Molecular cell* **35**, 716-723, doi:10.1016/j.molcel.2009.06.034 (2009).

196 Enoiu, M., Jiricny, J. & Scharer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic acids research* **40**, 8953-8964, doi:10.1093/nar/gks670 (2012).

197 Wang, X. *et al.* Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair. *Molecular and cellular biology* **21**, 713-720, doi:10.1128/MCB.21.3.713-720.2001 (2001).

198 Moldovan, G. L. & D'Andrea, A. D. How the fanconi anemia pathway guards the genome. *Annual review of genetics* **43**, 223-249, doi:10.1146/annurev-genet-102108-134222 (2009).

199 Kee, Y. & D'Andrea, A. D. Expanded roles of the Fanconi anemia pathway in preserving genomic stability. *Genes & development* **24**, 1680-1694, doi:10.1101/gad.1955310 (2010).

200 Ciccia, A. *et al.* Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Molecular cell* **25**, 331-343, doi:10.1016/j.molcel.2007.01.003 (2007).

201 Collins, N. B. *et al.* ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function. *Blood* **113**, 2181-2190, doi:10.1182/blood-2008-05-154294 (2009).

202 Garcia-Higuera, I. *et al.* Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Molecular cell* **7**, 249-262 (2001).

203 Huang, M. *et al.* Human MutS and FANCM complexes function as redundant DNA damage sensors in the Fanconi Anemia pathway. *DNA repair* **10**, 1203-1212, doi:10.1016/j.dnarep.2011.09.006 (2011).

204 Zhang, J. & Walter, J. C. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA repair* **19**, 135-142, doi:10.1016/j.dnarep.2014.03.018 (2014).

205 Nojima, K. *et al.* Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer research* **65**, 11704-11711, doi:10.1158/0008-5472.CAN-05-1214 (2005).

206 Gan, G. N., Wittschieben, J. P., Wittschieben, B. O. & Wood, R. D. DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res* **18**, 174-183, doi:10.1038/cr.2007.117 (2008).

207 Sharma, S., Helchowski, C. M. & Canman, C. E. The roles of DNA polymerase zeta and the Y family DNA polymerases in promoting or preventing genome instability. *Mutation research* **743-744**, 97-110, doi:10.1016/j.mrfmmm.2012.11.002 (2013).

208 Minko, I. G. *et al.* Replication bypass of the acrolein-mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family. *Chem Res Toxicol* **21**, 1983-1990, doi:10.1021/tx800174a (2008).

209 Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129**, 665-679, doi:10.1016/j.cell.2007.05.003 (2007).

210 Ulrich, H. D. Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA repair* **8**, 461-469, doi:10.1016/j.dnarep.2009.01.006 (2009).

211 Wang, L. C. & Gautier, J. The Fanconi anemia pathway and ICL repair: implications for cancer therapy. *Critical reviews in biochemistry and molecular biology* **45**, 424-439, doi:10.3109/10409238.2010.502166 (2010).

212 Wang, A. T. & Smogorzewska, A. SnapShot: Fanconi Anemia and Associated Proteins. *Cell* **160**, 354-354 e351, doi:10.1016/j.cell.2014.12.031 (2015).

213 de Laat, W. L., Jaspers, N. G. & Hoeijmakers, J. H. Molecular mechanism of nucleotide excision repair. *Genes & development* **13**, 768-785 (1999).
214 Thoms, K. M., Kuschal, C. & Emmert, S. Lessons learned from DNA repair defective syndromes. *Exp Dermatol* **16**, 532-544, doi:10.1111/j.1600-0625.2007.00559.x (2007).
215 Gillet, L. C. & Scharer, O. D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* **106**, 253-276, doi:10.1021/cr040483f (2006).
216 Chang, D. J. & Cimprich, K. A. DNA damage tolerance: when it's OK to make mistakes. *Nature chemical biology* **5**, 82-90, doi:10.1038/nchembio.139 (2009).
217 Haynes, B., Saadat, N., Myung, B. & Shekhar, M. P. Crosstalk between translesion synthesis, Fanconi anemia network, and homologous recombination repair pathways in interstrand DNA crosslink repair and development of chemoresistance. *Mutat Res Rev Mutat Res* **763**, 258-266, doi:10.1016/j.mrrev.2014.11.005 (2015).
218 Ghosal, G. & Chen, J. DNA damage tolerance: a double-edged sword guarding the genome. *Translational cancer research* **2**, 107-129, doi:10.3978/j.issn.2218-676X.2013.04.01 (2013).
219 Fujiwara, Y. *et al.* Xeroderma pigmentosum groups C and F: additional assignments and a review of the subjects in Japan. *Journal of radiation research* **26**, 443-449 (1985).
220 Niedernhofer, L. J. *et al.* A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**, 1038-1043, doi:10.1038/nature05456 (2006).
221 Overmeer, R. M. *et al.* Replication protein A safeguards genome integrity by controlling NER incision events. *J Cell Biol* **192**, 401-415, doi:10.1083/jcb.201006011 (2011).
222 Geng, L., Huntoon, C. J. & Karnitz, L. M. RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network. *J Cell Biol* **191**, 249-257, doi:10.1083/jcb.201005101 (2010).
223 Fu, D. *et al.* Recruitment of DNA polymerase eta by FANCD2 in the early response to DNA damage. *Cell cycle* **12**, 803-809, doi:10.4161/cc.23755 (2013).
224 Zhang, J., Zhao, D., Wang, H., Lin, C. J. & Fei, P. FANCD2 monoubiquitination provides a link between the HHR6 and FA-BRCA pathways. *Cell cycle* **7**, 407-413 (2008).
225 Davies, A. A., Huttner, D., Daigaku, Y., Chen, S. & Ulrich, H. D. Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein a. *Molecular cell* **29**, 625-636, doi:10.1016/j.molcel.2007.12.016 (2008).
226 Kim, H., Yang, K., Dejsuphong, D. & D'Andrea, A. D. Regulation of Rev1 by the Fanconi anemia core complex. *Nature structural & molecular biology* **19**, 164-170, doi:10.1038/nsmb.2222 (2012).
227 Huang, T. T. *et al.* Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nat Cell Biol* **8**, 339-347, doi:10.1038/ncb1378 (2006).
228 Nakanishi, K. *et al.* Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1110-1115, doi:10.1073/pnas.0407796102 (2005).
229 Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Molecular cell* **39**, 36-47, doi:10.1016/j.molcel.2010.06.023 (2010).
230 Yamamoto, K. *et al.* Fanconi anemia protein FANCD2 promotes immunoglobulin gene conversion and DNA repair through a mechanism related to homologous recombination. *Molecular and cellular biology* **25**, 34-43, doi:10.1128/MCB.25.1.34-43.2005 (2005).
231 Zhang, N., Liu, X., Li, L. & Legerski, R. Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. *DNA repair* **6**, 1670-1678, doi:10.1016/j.dnarep.2007.06.002 (2007).
232 Nakanishi, K. *et al.* Homology-directed Fanconi anemia pathway cross-link repair is dependent on DNA replication. *Nature structural & molecular biology* **18**, 500-503, doi:10.1038/nsmb.2029 (2011).
233 Adamo, A. *et al.* Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Molecular cell* **39**, 25-35, doi:10.1016/j.molcel.2010.06.026 (2010).

234 Pace, P. *et al.* Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. *Science* **329**, 219-223, doi:10.1126/science.1192277 (2010).
235 Bunting, S. F. *et al.* BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. *Molecular cell* **46**, 125-135, doi:10.1016/j.molcel.2012.02.015 (2012).
236 Murina, O. *et al.* FANCD2 and CtIP cooperate to repair DNA interstrand crosslinks. *Cell reports* **7**, 1030-1038, doi:10.1016/j.celrep.2014.03.069 (2014).
237 Unno, J. *et al.* FANCD2 binds CtIP and regulates DNA-end resection during DNA interstrand crosslink repair. *Cell reports* **7**, 1039-1047, doi:10.1016/j.celrep.2014.04.005 (2014).
238 Huang, M. *et al.* The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. *Molecular cell* **39**, 259-268, doi:10.1016/j.molcel.2010.07.005 (2010).
239 Yamamoto, K. *et al.* Fanconi anemia FANCG protein in mitigating radiation- and enzyme-induced DNA double-strand breaks by homologous recombination in vertebrate cells. *Molecular and cellular biology* **23**, 5421-5430 (2003).
240 Howard, S. M., Yanez, D. A. & Stark, J. M. DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. *PLoS genetics* **11**, e1004943, doi:10.1371/journal.pgen.1004943 (2015).
241 Callen, E. *et al.* A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood* **105**, 1946-1949, doi:10.1182/blood-2004-07-2588 (2005).
242 Schneider, M., Chandler, K., Tischkowitz, M. & Meyer, S. Fanconi anaemia: genetics, molecular biology, and cancer - implications for clinical management in children and adults. *Clinical genetics* **88**, 13-24, doi:10.1111/cge.12517 (2015).
243 Kelly, P. F. *et al.* Stem cell collection and gene transfer in Fanconi anemia. *Molecular therapy : the journal of the American Society of Gene Therapy* **15**, 211-219, doi:10.1038/sj.mt.6300033 (2007).
244 Ceccaldi, R. *et al.* Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell* **11**, 36-49, doi:10.1016/j.stem.2012.05.013 (2012).
245 Tulpule, A. *et al.* Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. *Blood* **115**, 3453-3462, doi:10.1182/blood-2009-10-246694 (2010).
246 Rosenberg, P. S., Alter, B. P. & Ebell, W. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica* **93**, 511-517, doi:10.3324/haematol.12234 (2008).
247 Alter, B. P., Greene, M. H., Velazquez, I. & Rosenberg, P. S. Cancer in Fanconi anemia. *Blood* **101**, 2072, doi:10.1182/blood-2002-11-3597 (2003).
248 Cioc, A. M., Wagner, J. E., MacMillan, M. L., DeFor, T. & Hirsch, B. Diagnosis of myelodysplastic syndrome among a cohort of 119 patients with fanconi anemia: morphologic and cytogenetic characteristics. *American journal of clinical pathology* **133**, 92-100, doi:10.1309/AJCP7W9VMJENZOVG (2010).
249 Gregory, J. J., Jr. *et al.* Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2532-2537, doi:10.1073/pnas.051609898 (2001).
250 Waisfisz, Q. *et al.* Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. *Nature genetics* **22**, 379-383, doi:10.1038/11956 (1999).
251 Soulier, J. *et al.* Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway. *Blood* **105**, 1329-1336, doi:10.1182/blood-2004-05-1852 (2005).

252 Ceccaldi, R. *et al.* Spontaneous abrogation of the G(2)DNA damage checkpoint has clinical benefits but promotes leukemogenesis in Fanconi anemia patients. *The Journal of clinical investigation* **121**, 184-194, doi:10.1172/JCI43836 (2011).

253 Auerbach, A. D., Rogatko, A. & Schroeder-Kurth, T. M. International Fanconi Anemia Registry: relation of clinical symptoms to diepoxybutane sensitivity. *Blood* **73**, 391-396 (1989).

254 Seyschab, H. *et al.* Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. *Blood* **85**, 2233-2237 (1995).

255 Neveling, K., Endt, D., Hoehn, H. & Schindler, D. Genotype-phenotype correlations in Fanconi anemia. *Mutation research* **668**, 73-91, doi:10.1016/j.mrfmmm.2009.05.006 (2009).

256 Saito, H., Hammond, A. T. & Moses, R. E. Hypersensitivity to oxygen is a uniform and secondary defect in Fanconi anemia cells. *Mutation research* **294**, 255-262 (1993).

257 Schindler, D. & Hoehn, H. Fanconi anemia mutation causes cellular susceptibility to ambient oxygen. *American journal of human genetics* **43**, 429-435 (1988).

258 Joksic, I. *et al.* Dysfunctional telomeres in primary cells from Fanconi anemia FANCD2 patients. *Genome integrity* **3**, 6, doi:10.1186/2041-9414-3-6 (2012).

259 Kirwan, M. & Dokal, I. Dyskeratosis congenita, stem cells and telomeres. *Biochim Biophys Acta* **1792**, 371-379, doi:10.1016/j.bbadis.2009.01.010 (2009).

260 Leteurtre, F. *et al.* Accelerated telomere shortening and telomerase activation in Fanconi's anaemia. *British journal of haematology* **105**, 883-893 (1999).

261 Callen, E. *et al.* Breaks at telomeres and TRF2-independent end fusions in Fanconi anemia. *Human molecular genetics* **11**, 439-444 (2002).

262 Dufour, C. *et al.* TNF-alpha and IFN-gamma are overexpressed in the bone marrow of Fanconi anemia patients and TNF-alpha suppresses erythropoiesis in vitro. *Blood* **102**, 2053-2059, doi:10.1182/blood-2003-01-0114 (2003).

263 Rosselli, F., Sanceau, J., Gluckman, E., Wietzerbin, J. & Moustacchi, E. Abnormal lymphokine production: a novel feature of the genetic disease Fanconi anemia. II. In vitro and in vivo spontaneous overproduction of tumor necrosis factor alpha. *Blood* **83**, 1216-1225 (1994).

264 Schultz, J. C. & Shahidi, N. T. Tumor necrosis factor-alpha overproduction in Fanconi's anemia. *American journal of hematology* **42**, 196-201 (1993).

265 Ibanez, A. *et al.* Elevated levels of IL-1beta in Fanconi anaemia group A patients due to a constitutively active phosphoinositide 3-kinase-Akt pathway are capable of promoting tumour cell proliferation. *The Biochemical journal* **422**, 161-170, doi:10.1042/BJ20082118 (2009).

266 Li, J. *et al.* TNF-alpha induces leukemic clonal evolution ex vivo in Fanconi anemia group C murine stem cells. *The Journal of clinical investigation* **117**, 3283-3295, doi:10.1172/JCI31772 (2007).

267 Wajant, H., Pfizenmaier, K. & Scheurich, P. Non-apoptotic Fas signaling. *Cytokine & growth factor reviews* **14**, 53-66 (2003).

268 Osborn, L., Kunkel, S. & Nabel, G. J. Tumor necrosis factor alpha and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor kappa B. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2336-2340 (1989).

269 Jacome, A. *et al.* A simplified approach to improve the efficiency and safety of ex vivo hematopoietic gene therapy in fanconi anemia patients. *Human gene therapy* **17**, 245-250, doi:10.1089/hum.2006.17.245 (2006).

270 Mehta, P. A. *et al.* Etanercept treatment in Fanconi anaemia; combined US and Italian experience. *British journal of haematology* **158**, 809-811, doi:10.1111/j.1365-2141.2012.09250.x (2012).

271 Zhang, X. *et al.* Defective homing is associated with altered Cdc42 activity in cells from patients with Fanconi anemia group A. *Blood* **112**, 1683-1686, doi:10.1182/blood-2008-03-147090 (2008).

272 Pilonetto, D. V. *et al.* FANCD2 Western blot as a diagnostic tool for Brazilian patients with Fanconi anemia. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.]* **42**, 237-243 (2009).

273 Vinciguerra, P., Godinho, S. A., Parmar, K., Pellman, D. & D'Andrea, A. D. Cytokinesis failure occurs in Fanconi anemia pathway-deficient murine and human bone marrow hematopoietic cells. *The Journal of clinical investigation* **120**, 3834-3842, doi:10.1172/JCI43391 (2010).

274 Schuler, D., Kiss, A. & Fabian, F. Chromosomal peculiarities and "in vitro" examinations in Fanconi's anaemia. *Humangenetik* **7**, 314-322 (1969).

275 Antonio Casado, J. *et al.* A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. *Journal of medical genetics* **44**, 241-249, doi:10.1136/jmg.2006.044719 (2007).

276 Stucki, A. *et al.* Decreased rejection and improved survival of first and second marrow transplants for severe aplastic anemia (a 26-year retrospective analysis). *Blood* **92**, 2742-2749 (1998).

277 Dufour, C. & Svahn, J. Fanconi anaemia: new strategies. *Bone marrow transplantation* **41 Suppl 2**, S90-95, doi:10.1038/bmt.2008.63 (2008).

278 Shimamura, A. & Alter, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood reviews* **24**, 101-122, doi:10.1016/j.blre.2010.03.002 (2010).

279 Velazquez, I. & Alter, B. P. Androgens and liver tumors: Fanconi's anemia and non-Fanconi's conditions. *American journal of hematology* **77**, 257-267, doi:10.1002/ajh.20183 (2004).

280 Scheckenbach, K. *et al.* Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood Cells Mol Dis* **48**, 128-131, doi:10.1016/j.bcmd.2011.11.006 (2012).

281 Guinan, E. C., Lopez, K. D., Huhn, R. D., Felser, J. M. & Nathan, D. G. Evaluation of granulocyte-macrophage colony-stimulating factor for treatment of pancytopenia in children with fanconi anemia. *J Pediatr* **124**, 144-150 (1994).

282 Rackoff, W. R. *et al.* Prolonged administration of granulocyte colony-stimulating factor (filgrastim) to patients with Fanconi anemia: a pilot study. *Blood* **88**, 1588-1593 (1996).

283 Gluckman, E., Berger, R. & Dutreix, J. Bone marrow transplantation for Fanconi anemia. *Semin Hematol* **21**, 20-26 (1984).

284 Zecca, M. *et al.* HLA-haploidentical T cell-depleted allogeneic hematopoietic stem cell transplantation in children with Fanconi anemia. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* **20**, 571-576, doi:10.1016/j.bbmt.2014.01.015 (2014).

285 MacMillan, M. L. *et al.* Alternative donor hematopoietic cell transplantation for Fanconi anemia. *Blood* **125**, 3798-3804, doi:10.1182/blood-2015-02-626002 (2015).

286 Alter, B. P., Giri, N., Pan, Y., Savage, S. A. & Pinto, L. A. Antibody response to human papillomavirus vaccine in subjects with inherited bone marrow failure syndromes. *Vaccine* **32**, 1169-1173, doi:10.1016/j.vaccine.2013.11.048 (2014).

287 Walsh, C. E. *et al.* Retroviral-mediated gene transfer for Fanconi anemia group A patients - a clinical trial. *Blood* **98**, 718a (2001).

288 Liu, J. M. *et al.* Engraftment of hematopoietic progenitor cells transduced with the Fanconi anemia group C gene (FANCC). *Human gene therapy* **10**, 2337-2346, doi:10.1089/10430349950016988 (1999).

289 Tolar, J. *et al.* Gene therapy for Fanconi anemia: one step closer to the clinic. *Human gene therapy* **23**, 141-144, doi:10.1089/hum.2011.237 (2012).

290 Gonzalez-Murillo, A. *et al.* Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. *Human gene therapy* **21**, 623-630, doi:10.1089/hum.2009.141 (2010).

291 Liu, G. H. *et al.* Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. *Nature communications* **5**, 4330, doi:10.1038/ncomms5330 (2014).

292 Osborn, M. J. *et al.* Fanconi anemia gene editing by the CRISPR/Cas9 system. *Human gene therapy* **26**, 114-126, doi:10.1089/hum.2014.111 (2015).

293 Hussain, T. & Mulherkar, R. Lymphoblastoid Cell lines: a Continuous in Vitro Source of Cells to Study Carcinogen Sensitivity and DNA Repair. *Int J Mol Cell Med* **1**, 75-87 (2012).

294 Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11-23 (2001).

295 Yanez-Munoz, R. J. *et al.* Effective gene therapy with nonintegrating lentiviral vectors. *Nature medicine* **12**, 348-353, doi:10.1038/nm1365 (2006).

296 Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature biotechnology* **22**, 589-594, doi:10.1038/nbt957 (2004).

297 Yagasaki, H. *et al.* A cytoplasmic serine protein kinase binds and may regulate the Fanconi anemia protein FANCA. *Blood* **98**, 3650-3657 (2001).

298 Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *Journal of virology* **72**, 8463-8471 (1998).

299 Burns, J. C., Friedmann, T., Driever, W., Burrascano, M. & Yee, J. K. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8033-8037 (1993).

300 Reiser, J. *et al.* Transduction of nondividing cells using pseudotyped defective high-titer HIV type 1 particles. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15266-15271 (1996).

301 Schenborn, E. & Groskreutz, D. Reporter gene vectors and assays. *Mol Biotechnol* **13**, 29-44, doi:10.1385/MB:13:1:29 (1999).

302 Charrier, S. *et al.* Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction. *Gene therapy* **18**, 479-487, doi:10.1038/gt.2010.163 (2011).

303 Romero, Z. *et al.* beta-globin gene transfer to human bone marrow for sickle cell disease. *The Journal of clinical investigation*, doi:10.1172/JCI67930 (2013).

304 Perumbeti, A. *et al.* A novel human gamma-globin gene vector for genetic correction of sickle cell anemia in a humanized sickle mouse model: critical determinants for successful correction. *Blood* **114**, 1174-1185, doi:10.1182/blood-2009-01-201863 (2009).

305 Smith, J., Tho, L. M., Xu, N. & Gillespie, D. A. The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res* **108**, 73-112, doi:10.1016/B978-0-12-380888-2.00003-0 (2010).

306 Lin, Y. & Waldman, A. S. Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA. *Nucleic acids research* **29**, 3975-3981 (2001).

307 Kass, E. M. & Jasin, M. Collaboration and competition between DNA double-strand break repair pathways. *FEBS letters* **584**, 3703-3708, doi:10.1016/j.febslet.2010.07.057 (2010).

308 van Nierop, G. P., de Vries, A. A., Holkers, M., Vrijsen, K. R. & Goncalves, M. A. Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic acids research* **37**, 5725-5736, doi:10.1093/nar/gkp643 (2009).

309 Wang, J. *et al.* Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome research* **22**, 1316-1326, doi:10.1101/gr.122879.111 (2012).

310 Li, H. *et al.* In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217-221, doi:10.1038/nature10177 (2011).

311 Rahman, S. H., Maeder, M. L., Joung, J. K. & Cathomen, T. Zinc-finger nucleases for somatic gene therapy: the next frontier. *Human gene therapy* **22**, 925-933, doi:10.1089/hum.2011.087 (2011).

312 Ousterout, D. G. *et al.* Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Molecular therapy : the journal of the American Society of Gene Therapy* **21**, 1718-1726, doi:10.1038/mt.2013.111 (2013).

313 Meier, D. & Schindler, D. Fanconi anemia core complex gene promoters harbor conserved transcription regulatory elements. *PloS one* **6**, e22911, doi:10.1371/journal.pone.0022911 (2011).

314 Saydaminova, K. *et al.* Efficient genome editing in hematopoietic stem cells with helper-dependent Ad5/35 vectors expressing site-specific endonucleases under microRNA regulation. *Mol Ther Methods Clin Dev* **1**, 14057, doi:10.1038/mtm.2014.57 (2015).

315 Liu, J., Gaj, T., Wallen, M. C. & Barbas, C. F., 3rd. Improved cell-penetrating zinc-finger nuclease proteins for precision genome engineering. *Molecular therapy. Nucleic acids* **4**, e232, doi:10.1038/mtna.2015.6 (2015).

316 Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells in vivo. *Blood* **117**, 4460-4466, doi:10.1182/blood-2010-08-303537 (2011).

317 Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology*, doi:10.1038/nbt.3290 (2015).

318 Nakamura-Ishizu, A., Takizawa, H. & Suda, T. The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development* **141**, 4656-4666, doi:10.1242/dev.106575 (2014).

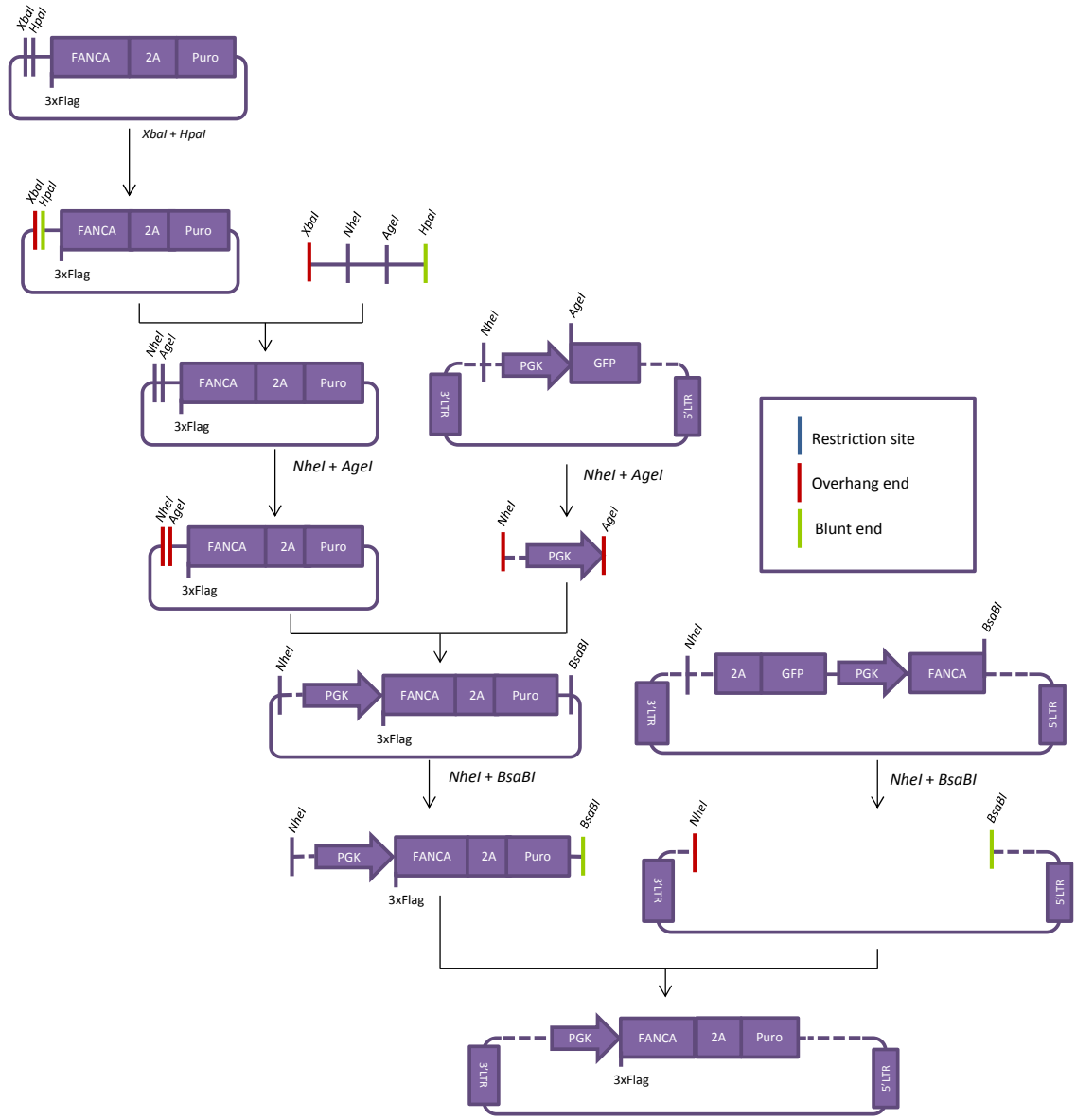
319 Wunderlich, M. *et al.* AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* **24**, 1785-1788, doi:10.1038/leu.2010.158 (2010).

320 Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. Systematic determination of the packaging limit of lentiviral vectors. *Human gene therapy* **12**, 1893-1905, doi:10.1089/104303401753153947 (2001).

321 Fingar, D. C. *et al.* mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E. *Molecular and cellular biology* **24**, 200-216 (2004).

APPENDIX I:

SUPPLEMENTARY FIGURES



Supplementary figure 1: Cloning strategy of FANCA-Puro donor.

APPENDIX II:

PUBLISHED ARTICLES RELATED TO THE PRESENT DOCTORAL THESIS